REPORT RESUMES

ED 003 271 24

EDUCABILITY OF STUDENTS IN THE VISUALIZATION OF OBJECTS IN SPACE.

BY- CARPENTER, FINLEY AND OTHERS
MICHIGAN UNIV., ANN ARBOR
REPORT NUMBER CRP-1474 PUB DATE 65
EDRS PRICE MF-$0.18 HC-$4.92 123P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, *APTITUDE, APTITUTE TESTS, PERFORMANCE, *LOW ACHIEVERS, *FILMS, FILMSTRIPS, TESTING, SECONDARY SCHOOL STUDENTS, GRADE 8, GRADE 9, SPACE RELATIONS TEST, DIFFERENTIAL APTITUDE BATTERY, ANN ARBOR, MICHIGAN, GEOMETRIC INVENTORY

AN ATTEMPT WAS MADE TO DETERMINE THE EFFECT OF SYSTEMATIC INSTRUCTION IN SPACE RELATIONS ON RESULTS OF TESTS DESIGNED TO MEASURE SPACE RELATIONS AS AN APTITUDE. OVER 800 STUDENTS, IN 4 SCHOOLS, WERE SELECTED FROM THE EIGHTH- AND NINTH-GRADE LEVELS TO PARTICIPATE IN THE STUDY. INSTRUCTIONAL TREATMENTS VARIED FROM MULTIPLE INPUTS TO ZERO INSOFAR AS THE TEACHING OF SPACE RELATIONS WAS CONCERNED. SPECIFICALLY, EIGHT TREATMENTS IDENTIFIED BY THE KIND AND NUMBER OF INSTRUCTIONAL INPUTS--CONTROL (NO INSTRUCTION), FILM, FILMSTRIP, AND PROGRAMED INSTRUCTION, PLUS THE COMBINATIONS OF FILM AND PROGRAMED INSTRUCTION, FILMSTRIP AND PROGRAMED INSTRUCTION, AND FILM, FILMSTRIP, PROGRAMED INSTRUCTION. TWO HYPOTHESES WERE EVALUATED--(1) INSTRUCTIONAL INPUTS THAT INTERVENE BETWEEN PRE- AND POST-TESTS OF SPACE RELATIONS WOULD HAVE NO SIGNIFICANT EFFECT ON RELIABILITY COEFFICIENTS WHEN COMPARED WITH THE CONTROL GROUP DATA AND (2) THE AMOUNT AND KIND OF INSTRUCTIONAL INPUTS WOULD HAVE NO EFFECT ON CHANGING MEAN SCORES ON CRITERION MEASURES. RESULTS INDICATED THAT NO CLEAR-CUT ASSESSMENT OF EITHER HYPOTHESIS COULD BE MADE. THE MOST CONVINCING OUTCOME WAS THAT STUDENTS WITH LOW I.Q.'S WHO HAD OFTEN TRANSFERRED FROM ONE SCHOOL TO ANOTHER GAINED MOST FROM PROGRAMED INSTRUCTION. THERE IS A DISCUSSION ON WAYS IN WHICH FUTURE ATTEMPTS TO STUDY THE CENTRAL PROBLEM MIGHT BE DIRECTED. (JC)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the
person or organization originating it. Points of view or opinions
stated do not necessarily represent official Office of Education
position or policy.

EDUCABILITY OF STUDENTS IN THE VISUALIZATION OF OBJECTS IN SPACE

Cooperative Research Project No. 1474

Finley Carpenter
Project Director and Principal Investigator

Erwin H. Brinkmann
Daniel S. Lirones
Research Associates

The University of Michigan
Ann Arbor, Michigan

1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ED 003271

PRECEDING PAGE BLANK- NOT FILMED

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF ILLUSTRATIONS ix

ABSTRACT xi

Chapter

1. INTRODUCTION 1
    - Problem 1
    - Signficance of the Problem 2
    - Hypotheses 3
    - Additional Objectives 4
    - Related Research 5
        - Studies Pertaining to Programmed Instruction 7

2. PROCEDURE 9
    - Development of Instructional Materials 9
    - The Motion Picture 9
    - Program Development 11
        - Behavioral Specification 12
        - Program Content 13
        - Program Organization Chart 13
        - Frame-Writing and Testing 15
        - Revised Program Design 17
    - The Film Strip 19
    - Tactual-Kinesthetic Materials 19
    - Sampling 21
        - Ypsilanti: East Junior High School 22
        - St. Clair Shores: Ottawa Junior High School 22
        - Walled Lake Junior High School 22
        - Saline High School 23
    - The Ypsilanti Pilot Study 23
    - Data Collection Procedures 23
    - Variables 24
        - Space Relations Test 24
        - Content-Learning Measure 25
        - Additional Measures 26
    - Summary of Pilot Procedures 27

TABLE OF CONTENTS (Continued)

Page

Procedure for the Experiment Itself 28
  St. Clair Shores 28
  Walled Lake 29
  Saline 30
Summary 30

3. RESULTS OF THE PILOT STUDY 33

Program Performance 33
  Time Data 33
  Error Rate 33
  Geometric Inventory Analysis 35
Findings Pertaining to Hypotheses 35
  Hypothesis 1 35
  Hypothesis 2 37
Additional Findings 37
  Differential Effects 37
  Relation to Residual Variables 41
  Attitude Questionnaire 44

4. COMBINED RESULTS 47

The First Hypothesis 47
The Second Hypothesis 53
  High and Low Intelligence 55
  Rank orders of Treatments 59

5. DISCUSSION 64

6. SUMMARY AND CONCLUSIONS 71

Conclusions 72

APPENDIX A. ORGANIZATIONAL STRUCTURE FOR PROGRAM DEVELOPMENT 74

A1. Approach to Programming Visual Space Perception 75

APPENDIX B. DIRECTIONS FOR ADMINISTRATION OF THE SPACE RELATIONS TEST 92

APPENDIX C. INSTRUCTIONS FOR GEOMETRIC INVENTORY 96

APPENDIX D. GEOMETRIC INVENTORY 98

APPENDIX E. DIRECTIONS FOR USING THE PROGRAMMED MATERIALS 102

TABLE OF CONTENTS (Concluded)

Page

APPENDIX F. RAW DATA FOR EXPERIMENTAL AND CONTROL GROUPS 104

REFERENCES 108

PRECEDING PAGE BLANK- NOT FILMED

LIST OF TABLES

Table		Page
1.	Program Composition	18
2.	Treatment by Groups for St. Clair Shores Samples	28
3.	Treatment by Groups of Walled Lake Samples, Eighth Grade	29
4.	Treatment by Groups of Walled Lake Samples, Ninth Grade	29
5.	Treatment by Groups for Saline Samples	30
6.	Experimental Design Showing Treatments by School, Grade, and Sex	31
7.	Time Data for Program, Units I-X, for 28 Eigth-Grade Students	34
8.	Error Data for Program, Units I-X, for 28 Eighth-Grade Students	34
9.	Pretest and Posttest Range, Mean, and Standard Deviation of Scores on the Geometric Inventory	36
10.	Comparison of the Experimental and Control Groups on the Geometric Inventory Pretest	36
11.	Comparison of the Experimental and Control Groups on the Geometric Inventory Posttest	36
12.	Correlation of Pretest and Posttest Space Relations Scores	38
13.	Equivalence of Groups on Precourse Measures Used as Basis for Matching Pairs	38
14.	A Comparison of the Experimental and Control Groups on Posttest Space Relations Scores	39
15.	Gains from Pretest to Posttest as Indicated by Percentile Norms for the Respective Forms and Sex	39
16.	Correlation Between Pretest and Posttest Sapce Relations Scores for Middle Half of Sample	41
17.	Correlation of Pre-Experimental Measures of Intelligence, Reading Comprehension, and Arithmetic Concepts with Experimentally Manipulated Variables for the Experimental Group	43

LIST OF TABLES (Concluded)

Table Page

18. Relationship Between Performance on the Posttest Geometric Inventory and Expressed Attitude to Statement, "Teachers Can Teach Much Better Than a Program" 45

19. Relationship Between Performance on the Posttest Geometric Inventory and Expressed Desire for Degree of Use of Programmed Instruction as an Instructional Input 45

20. Correlation Between Pre and Post Space Relations Test by Treatment (Sex = Male) 48

21. Correlation Between Pre and Post Space Relations Test by Treatment (Sex = Female) 49

22. Correlations and z-Transformations under Pooled Groups by Treatment and by Sex 50

23. Significant Differences and Levels Between Correlations of Groups Pooled by Treatment, Sex and School. (Space Relations Test) 51

24. Adjusted Mean Scores by Covariance Analysis for Males by Treatment within Schools (Space Relations Test) 54

25. Adjusted Mean Scores by Covariance Analysis for Females by Treatment within Schools (Space Relations Test) 56

26. Covariance Analysis of Pre- and Post-Space Relations Test Scores for Subjects Grouped as High MA (Mean I.Q. 110) 57

27. Covariance Analysis of Pre- and Post-Space Relations Test Scores for Subjects Grouped as Low MA (Mean I.Q. 94) 58

28. Expected Versus Obtained Rank Orders of Gains by Treatment (Gains on Basis of Adjusted Means) 59

29. Adjusted Mean Schores by Covariance Analysis for Males by Treatment within Schools (Geometric Inventory Test) 62

30. Adjusted Mean Scores by Covariance Analysis for Females by Treatment within Schools (Geometric Inventory Test) 63

LIST OF ILLUSTRATIONS

Plate | Page

I-II. The Solids in the Kit 20

Figure

1. A Chart to Schematize Relationships for Estimation of Length 14

2. Pretest-Posttest Change for Each Quarter of the Experimental and Control Groups on the Space Relations Test 40

3. Pretest-Posttest Change for Each Quarter of the Experimental Group on the Geometric Inventory 42

4. Adjusted Mean Gains by Treatments and by Pooled Subjects 60

PRECEDING PAGE BLANK- NOT FILMED

ABSTRACT

The purpose of this study was to determine the effect of systematic instruction in space relations on results of tests designed to measure space relations as an aptitude. An initial consideration that led to the development of the project was the apparent disparity between the general intent of education and the manner in which measures of intellectual factors are often interpreted. The basic purpose of education is to identify and to rectify, insofar as possible, any deficiency that the student exhibits in terms of desirable knowledge and skills contained within the educational objectives. If measures of aptitude must be interpreted as limitations which nature has imposed upon the individual, decision-making in education that depends upon such information amounts large to the sorting of students to fit the demands of certain courses. If, however, such measures can be considerably altered through systematic instruction, scores on aptitude tests should perhaps function more to define specific problems than to be regarded as unalterable limitations. It is believed, therefore, that one of the major tasks that confront education is to determine what aspects of behavior germane to education can be influenced appreciably and which ones are relatively stable, that is, tend to resist attempts at modification.

Over 800 students in four schools from different communities in or near to the Detroit Metropolitan Area were selected from the eighth and ninth grades to participate in the study. Instructional treatments varied from multiple inputs to zero insofar as the teaching of space relations was concerned.

More specifically, eight different treatments were given, including the control group, which received no special instruction. The components used for teaching were: (a) a 505-frame program designed to help students develop flat patterns into solids. In conjunction with the program a tactual-kinesthetic kit of objects were used so that students could actually manipulate both flat patterns and solids to get the "feel" of the process called "development." As the program progressed, the student was called upon to depend more and more upon mental manipulation and less and less of concrete manipulation. (b) A 15-minute film in color was produced for the purpose of maximizing student motivation. It contained very little specific content on how to deal with problems of space perception, but it did provide a sample of practial activities that involved skill in space perception. The intent was to use the film so that students would regard space perception as an extremely desirable skill. The film also emphasized that the ability could be learned. (c) A filmstrip of 81 frames was constructed to illustrate the development of various solids from flat patterns. A sound tape was synchronized with the presentation of the frames.

ABSTRACT (Concluded)

The eight treatments were identified by the kind and number of instructional inputs, namely, as: control (no instruction), film, filmstrip, program, plus these combinations: film and program, filmstrip and program, and film, filmstrip, program.

Two hypotheses were evaluated: (a) instructional inputs that intervene between pre- and posttests of space relations (DAT) would have no significant effect on reliability coefficients when compared with the control group data, (b) the amount and kind of instructional inputs would have no appreciable effect on changing mean scores on criterion measures.

Results indicated no clear-cut assessment of either hypothesis could be made. The most convincing outcome involved data that showed students with low I.Q.s and who had transferred schools three times or more during their school years gained most from programmed instruction. When low I.Q. transients who had taken the program were compared with low I.Q. transients in the control groups, the former gained nearly 12 times more than the latter. These results suggest that for the transient student (as defined in this report) with low I.Q. materials in programmed form may be of special benefit.

In relation to the first hypothesis, where it was expected the instruction would lower reliability coefficients significantly when compared with control group data, 10 of the possible 28 comparisons showed that instructed groups had significantly lower pre- posttest reliability coefficients than the control groups. That finding suggests that in some groups initial rank orders of students shifted significantly on the posttest.

Results pertaining to the second hypothesis indicated that groups which received the program did much better than control groups on the posttest. That finding was true for pooled group results. But when groups were treated individually within schools, results seemed less significant. Again, the findings are not clear-cut.

Several artifacts and practial limitations probably combined to lessen the strength of the independent variables. In the discussion chapter these possible weaknesses were treated. Such treatment implies in what ways future attempts to examine the central problem may be improved over the operations in the present study.

CHAPTER ONE: INTRODUCTION

PROBLEM

The ability to manipulate geometric forms mentally is important in mathematics, the plastic arts, and certain phases of engineering. Aptitude scores on tests of spatial relations predict success in a variety of courses. While correlations are not extremely high, the tests are usually good predictors for groups past the age of sixteen. Among intelligence factors, perception of space relations is unique and one of the most stable.

Sometimes aptitude scores are interpreted to indicate the existence of inner capacities which determine both individual scores and the correlated success in learning. At other times, aptitude measures are said to point up only a set of symptoms predictive of achievement. The first interpretation rests on the assumption that the strength of genetic determiners can be measured indirectly by aptitude tests. The second interpretation involves the assumption that all persons measured have had about an equal opportunity to acquire the appropriate skills. Individual differences reflected by scores are said to measure the learning efficiency established by the individual and can be taken as characteristic of future success in learning.

The notion that aptitude tests measure genetic determiners indirectly can be challenged on the ground that no psychological traits have yet been traced to the genes in any definitive way. The claim involves mostly speculation and little if any solid evidence. Until the limits of stimulus control have been established for complex learning, there is probably no advantage in claiming that aptitude scores measure the effects of genetic determiners.

The assumption that all examinees have had about the same opportunity to acquire the skills necessary for answering test items correctly also appears tenuous. An "opportunity" is not wholly an objective fact; it is related to the status of the person, involving his health, attitudes, interests, motivation, and special habit patterns. Even siblings in the same home rarely have equal psychological environments. Different psychological environments have not been shown to produce highly similar modes in taking advantage of opportunity. Professor Skinner of Harvard has shown that providing a pigeon with an opportunity to do something atypical (such as bowl a ball down a miniature alleyway) is not sufficient to guarantee that he will act on the opportunity. Yet, when the immediate environment is properly managed, the probability approaches certainty that the pigeon will learn to "bowl" the ball. Efficient education does not seem simply a matter of opportunity.

Let effective education depend, tentatively, upon three sets of factors: (a) genetic endowment, (b) opportunity, defined as the conditions necessary to allow expression of desired behavior, and (c) management of conditions to implement opportunity. (Opportunity and management differ in that the latter covers the way conditions are used to take advantage of opportunity.) If only the first two sets of determiners are taken as necessary and sufficient, while ignoring the third set, then the following statements appear tenable:

(a) The role of the teacher is largely to provide maximum opportunity for the pupil to develop according to his natural endowment.

(b) Measures of capacity and achievement should be dependable predictors of success in any endeavor which requires those factors. Also, the accuracy of prediction should remain relatively constant across the range of management processes in education.

(c) The rank order of individual pretest aptitude scores interpreted to measure genetic and/or stable symptoms of achievement would not change appreciably after optimal management of learning involving the given aptitude.

This research was aimed primarily at evaluating statement "c," above. The following tasks point up the problem as two specific goals:

1. To determine whether or not special training given to students in dealing with problems of spatial relationships would alter the predictive power of the standardized criterion test.

2. To determine the effect of different amounts of instructional input on changes in scores on the standardized aptitude test in space relations and on a special achievement test designed to measure learning as the result of training.

SIGNIFICANCE OF THE PROBLEM

The principal objective of education is to provide and to use the most effective means for the development of human potential. It is perhaps safe to say that the maximum possible limits of the objective have not been reached. Among the efforts relevant to determining such limits must be included attempts to identify those traits which are educable and to discriminate them from those that are resistant to the influences of education.

The fact that most curricular experiences are organized on the basis of traditional subject matter rather than on the basis of a functional integration of desirable modes of behavior has probably served to inhibit an effective search for educable traits. A second likely inhibitor lies in the traditional perspective of intelligence and aptitudes. The presumed factors of intelligence have usually been regarded as somewhat resistant to education. Such orientation has emphasized the search for individual parameters as the bases for predicting success under the status quo. To what extent the presumed parameters are artifacts of the prevailing conditons of cultural nurture is a moot point. The concept of culture-free or culture-fair tests is relevant to the issue which centers on how much the forces of nurture contribute to human development.

Although the problem of nature versus nurture is perhaps insoluble in its commonly expressed form, it does imply a family of researchable tasks valuable to education. One of the most promising recent movements in education is identified in the objectives and procedures of those engaged in educational technology. The most characteristic features of that technology lie in efforts to cast educational objectives in the form of specific behaviors, to restructure content to facilitate its grasp by students, to use all available means of instruction, and to manage the instructional components to achieve maximum output per unit of input (effort by teacher and student). In short, educational technology is aimed at the maximum increase of effectiveness of education. The ability to determine those traits that are educable is fundamental to the success of the movement.

The present research is limited to exploring the educability of the intellectual trait known as the ability to deal with problems of space relations. A number of standardized tests have been developed to measure the space relations factor. Such tests are important in educational and vocational counseling. Students with low scores on space relations tests are not encouraged to pursue courses which apparently require a high loading of that factor to reach passable standards of achievement.

The fact that there is no systematic effort in the curriculum to increase students' skills in handling space perception problems means that training in the skill is left to chance. Abilities considered important for success in certain courses have often been regarded as intellectual prerequisites which somehow develop independently of formal education. When educable traits are left to develop through fortuitous experience they may be difficult to differentiate from intellectual dimensions which are not developed by education. Consequently, it is important to determine to what extent intellectual factors can be influenced by systematic instruction. Reliable information about such factors could provide the basis for a fruitful expansion of school curricula.

HYPOTHESES

The study was designed to collect data for the assessment of the following hypotheses:

1. Correlations between pre- and posttests of space relations (Differential Aptitude Test Battery) will not change significantly when space relations instruction intervenes between the two tests. The basis for making the statistical comparison will be the reliability coefficients taken from pre- and posttest scores of control groups.

2. There will be no significant differences in means on posttest criterion scores among groups with different amounts of training.

The first hypothesis is aimed at the reliability of the criterion measure when instruction intervenes between pre- and posttests. If no significant decrease in the reliability coefficient can be ascribed to training, the first hypothesis cannot be rejected. If, however, data can be used to reject the hypothesis the prevailing interpretation of space relations aptitude scores would be open to modification. A significant reduction in the reliability index would mean that training was the likely variable that shifted the rank orders of students from their pretest standings.

Rejection of the second hypothesis would mean that at least some of the instructional components influenced students to acquire skill in dealing with problems of space relations.

ADDITIONAL OBJECTIVES

In connection with the stated hypotheses special objectives merit recognition.

1. To develop educational materials for teaching spatial perception. Materials developed were in programmed form to include visual and tactual-kinesthetic components. More specifically the objective was to develop a linear program in booklet form and a kit of tactual-kinesthetic objects to use with the program. Other instructional materials developed included a film strip and a 15-minute color motion picture with sound.

2. To determine the efficiency of programmed instruction in teaching space relationships and to measure the influence of other special materials mentioned above.

3. To explore the effectiveness of the program at various ages. (This objective was necessarily narrowed to high school because standardized tests on space relations are inapplicable to the lower grades.)

RELATED RESEARCH

Because several references cover the history of programmed instruction (Galanter, 1959; Lumsdaine and Glaser, 1960; Smith and Moore, 1962), it is sufficient to note here that the fundamental idea of organizing subject matter into a sequence of easily understood, interrelated steps, progressing from simple to complex, was thoroughly emphasized and generally followed in developing the program for this investigation.

The direct training of some of the assumed components of visual space perception is described in a series of classic experiments (Thorndike and Woodworth, 1901) which attempted to determine the influence of prior training exercises in estimation on later abilities to estimate magnitudes of the same general sort, such as lengths and areas, differing in amount and in accessory qualities (shape, color, form). Although there was more improvement when the areas tested were similar in shape to those of the training series, the authors stated that "when one undergoes training in estimating certain magnitudes, he may improve in estimating others from various causes" (p. 394). They attributed improvement to the development of "more accurate mental standards...habits of making judgments in better ways, of making allowance for constant errors, and of avoiding certain prejudices" (p. 395).

Several studies illustrate a somewhat less direct training approach. In Van Voorhis' (1941) thesis, college students' training included estimating linear extent, angles, and areas; three-dimensional tick-tack-toe; various visualizing exercises; and practice with a stereoscope. Scores on a criterion measure (the "Cards and Figures" section of the Thurstone Test for Primary Mental Abilities) improved significantly compared with a control group. Brown (1954) compared spatial visualization (measured by the Space Relations Test of the Differential Aptitude Tests) of an experimental group whose training consisted of one year of plane geometry fused with chosen elements (58 theorems considered useful in developing visualization) of solid geometry with a control group at two levels of training: (1) students who completed the regular one-year plane geometry course, and (2) comparable students who completed a two-year sequence of plane geometry, advanced algebra, and solid geometry. The gain of the experimental group was less in every analysis. By way of explanation Brown (1954) suggested that the experimental group may have worked more slowly and more cautiously on the criterion measure than did the control group.

It is sometimes suggested that Ss with verbal pretraining learned to look for identifying parts of the stimuli. In experiments to determine the effect of verbal pretraining on discrimination performances, Gagne and Baker (1950) reported positive transfer after 32 training trials but not after 8 and 16 trials of predifferentiation training. Results of the study led to the suggestion that a minimum level of learning is required before transfer may be expected to occur. A similar conclusion could be drawn from Arnoult's (1953) study in which Ss failed to achieve positive transfer in perceiving various complex images. Failure to achieve positive transfer could be attributed to several factors: (1) insufficient levels of training, or (2) the absence of factors present in other studies showing positive transfer effects, for example, transfer of elements not specific to the stimulus material but specific to the experimental tasks or procedures.

One may ask whether learning tasks in the usual school curricula affect visual space perception. Several reports on effects of courses in vocational machine shop, descriptive geometry, and mechanical drawing on subsequent spatial relations test scores were considered relevant to this investiagation. A study conducted at the high school level (Mendicino, 1958) indicated that one school year in vocational machine shop and mechanical drawing had no more effect on the space perception test scores of tenth-grade boys than did a similar period in a nonvocational curriculum.

Another investigator (Myers, 1958) examined possible effects of mechanical drawing taken during high school on spatial relations tests administered at the Naval Academy. After comparing spatial test scores of cadets who had taken mechanical drawing with those who had not, and finding, on the average, equivalent scores, Myers concluded that previous training had very little effect on spatial relations test scores.

Is it possible that a different kind of emphasis in teaching could have altered the results? Reports of an average gain in spatial test scores of nearly one standard deviation for a group of 89 engineering students in their freshman year of college (Blade and Watson, 1955) encourage such speculation. The same investigators observed similar gains by 593 cadets at West Point and 114 engineering freshmen at the University of Wisconsin. Comparable nonengineering students measured at the same point in time had lower mean scores than engineering students and made gains over their entrance scores of only half a standard deviation. The difference in the gains between engineering students and nonengineering students was statistically significant. As a result of their study, Blade and Watson (1955) suggested the possible need to aid those with undeveloped spatial visualization by providing maximum opportunity to develop this capacity.

Other investigations using college students have reported similar gains. Comparison of pre- and posttest scores on a spatial relations test with a course in military topography and graphics as the experimental variable showed an increase of almost one standard deviation (Myers, 1951). Similarly, Meyers (1953) reported significant gains in scores on a spatial relations posttest following a year's training at West Point in descriptive geometry and engineering drawing.

The apparent contrast in effect of training at the high school and college level could conceivably be attributed to the greater spatial ability demands made by advanced courses. Beginning courses usually stress motor skills and the learning of symbols, thereby obscuring the relationships present in later stages when greater demands may be made on visualization. It would, of course, be presumptuous to ignore maturation as a contributing influence. How the test variables may further confound the results has been indicated in a study (Krumboltz and Chistal, 1960) designed to investigate the problem of practice effects in tests of spatial ability. They found that after a short interval (seven hours) practice effects were almost as high on an alternate form as when the identical test was readministered. As a result, the experimenters suggested using alternate types rather than alternate forms of spatial tests when short intervals are involved. Such results emphasize the necessity for adequate controls in the experimental design even when longer intervals are involved.

Studies Pertaining to Programmed Instruction

Reports on automated teaching or programmed instruction in schools, colleges, the Armed Forces (Lumsdaine and Glaser, 1960), and in industry (Hughes and McNamara, 1961), have suggested that this technique has considerable promise in reducing training time. Skill learning can be successfully accomplished through programmed instruction, as indicated by results obtained using the medium with a variety of content. The use of programmed texts in a series of high school mathematics couses (TEMAC, 1961) has led the investigators to conclude that programmed instruction resulted in marked savings in time and in improved effectiveness of learning, particularly in conjunction with a sympathetic teacher. Similar results have been produced in other content areas such as spelling (Moore and Smith, 1961), where sixth-grade pupils, using both the programmed text and machine formats, learned spelling to the same level as traditionally taught students "with a saving of at least fifty per cent time for all students" and even greater savings for the bright students.

On the college level entering freshmen have been taught how to use a university library (Wendt and Rust, 1962) by means of a branching program employing both pictorial and performance frames. In the Hamilton College Project (Blyth, 1961) experiments were conducted to determine whether the principles of programmed instruction could be applied successfully at practical levels to teach regular courses in the curriculum. After developing programs for French, German, logic, mathematics, and psychology, preliminary studies suggested that learning outcomes would be at least comparable to those achieved by more traditional techniques.

In a well-controlled investigation of effectiveness of a programmed text in medical parasitology, Green, Weiss, and Nice (1962) discovered that learning efficiency was considerably greater for almost all students than it was with nonprogrammed materials. The more productive students achieved the same high level in less time using programmed instruction, whereas the less productive students benefitted not only by spending less time on programmed content than they did on the conventional content, but they also achieved significantly higher scores on the examination.

The development of skills by programmed instruction in nonschool settings has also been reported. Trainees in a 7070 Data Processing System servicing courses using programmed texts showed significant gains in achievement and reduction of training time compared to control classes taught by the lecture-discussion method (Hughes and McNamara, 1961). Recruit firemen, after studying a 252-step lesson in "Fire Behavior" (Silvern and Fuchs, 1962), more than doubled their pretest scores on a written comprehension test. Military personnel, using filmed programmed materials in the Autotutor to learn fundamentals of electronics, did as well on the regular course examinations as did conventionally taught (lecture-discussion) control groups (Crowder and Zachert, 1960).

Using programming techniques, although not a verbal program, Hively (1962) trained preschool and first-grade children to make a series of progressively difficult visual discriminations. He found learning more efficient both in terms of error rate and time when discriminations were programmed from simple to complex versus training in the final discriminations alone.

These experimental studies indicate the variety of tasks for which programmed instrution has been effectively used and suggest the feasibility of using this instructional technique for additional training tasks.

CHAPTER TWO: PROCEDURE

DEVELOPMENT OF INSTRUCTIONAL MATERIALS

The treatment of experimental groups was preceded by the construction of four instructional components: (a) a 15-minute motion picture in color with sound, (b) a linear program of 505 frames, (c) a kit of tactual-kinesthetic solid objects and flat patterns used in conjunction with the program, and (d) a film strip of 81 frames with a synchronized tape recording.

THE MOTION PICTURE

Production and use of the film was for the purpose of influencing motivation. It represented an effort to depict a variety of scenes in which the skill of visualizing objects in space instrumented a number of successful experiences in everyday life. Its title, More Than Meets the Eye, was treated in the script to suggest an acquired ability to manipulate objects in space mentally for achieving desirable ends.

The rationale for choosing the film as a motivation input involved the promising use of its unique properties for stimulating orientation and imitation. Features peculiar to the motion picture as an instructional device include: (a) the stimulus similarity between pictured events and reality, (b) the storage and reproduction of a wide variety of visual episodes, (c) the flexibility offered for controlling visual perspective, and (d) effective controls for arresting and maintaining attention. Because the motion picture typically is shown in a darkened room, the vividness of competing stimuli is considerably reduced, while the critical stimuli are stressed by focusing them on a restricted area that is lighted sufficiently to make the details readily discernible. How are these unique properties used to stimulate effective orientation and imitation? And what does the result have to do with motivation? A quotation from Carpenter and Haddan (1964) points up the answer.

> Orientation. When a learner is oriented he has acquired a space-time predisposition, meaning that he has become aware of how to direct his behavior from a given point within a context toward likely reinforcing consequences. It also means that he has become aware of the temporal order of responses that result in reinforcement. The film can depict the space-time structure by presenting either a symbolic or actual picture of the context. Thus, discriminable stimuli associated with response chains can be indicated. Orientation is successful to the extent that cueing in the film provides a reliable map of reinforcing contingencies of similar situation later experienced by the student.

> Imitation. Orientation and imitation are similar in their overall effects. While orientation is a space-time map of reinforcing contingencies, imitation emphasizes the response side. The one being imitated (imitatee) demonstrates a mode of action that results in reinforcement. When similar behavior on the part of the imitator also bring reinforcement, the strength of imitation increases. While orientation requires no model of behavior, imitation depends upon actual demonstration of a behavioral pattern as a model to be approximated. Orientation can be provided symbolically while imitation is stimulated through demonstration.
>
> The net result of orientation and imitation is to increase the probability of certain behavior before the learner has actually behaved in ways indicated in the given orientation and the given model for imitation. Thus, imitation models and orientation act to influence the learner to want to act in a given manner. And this wanting to the the meaning of motivation.

An attempt was made to produce a motion picture that would promise to motivate students as pointed up in the above rationale. More specifically, however, the film was structured in accordance with the following: (quoted from Carpenter and Haddan, 1964).

> The influence of effective orientation is manifest in generalization or induction, meaning that the filmed and real contents, in which the learner operates, must contain common elements sufficient to predispose the desired behavior. Also, intended reinforcers must actually represent reinforcers to the learner. And so the requirements of an effective orientation include (a) a map that contains the essential structure and relations of some field, area, or context and (b) an indication of how a behaving organism functions within the context to bring about reinforcing consequences...A context having clearly defined contingencies of reinforcement functions somewhat as a maze in which blind alleys represent negative contingencies, and open pathways to the goal represent positive ones. A good orientation is a perspective of the real maze.
>
> A general formula for inducing imitation should probably contain (a) a demonstration of behavior which the observer is capable of reproducing, (b) clearly defined reinforcers which are contingent upon certain acts, (c) a context which is accessible to the learner, and (d) a type of demonstrator who is not aversive to the viewer - one that displays features that the observer likes.

During the planning stage of the film, the director and his research assistants met with the local producer of educational films at The University of Michigan. Several conferences were held in an effort to communicate the objective of the film and the kinds of events to be pictured. A professional script writer attended some of the conferences, during which the writer was informed of the purpose and specific properties sought. The writer then developed a script as the basis for filming.

Students selected for acting in the film were of ages comparable to those in the experimental sample. They were shown doing things that involved skill in space perception. Among the number of pictured episodes were scenes involving sports, hobbies, and vocational activities. The latter necessarily included a limited number of adults. One scene intended to further motivation and imitation was the showing of a class actually engaged in the use of the program. The entire effort was focused on situations typical of experiences in the lives of most eighth and ninth graders and emphasis was placed on reinforcement resulting from skill in space perception. (A copy of the film which accompanies this report is the final product.)

PROGRAM DEVELOPMENT

If programmed instruction is to be utilized as an instructional technique, an adequate program must first be developed. Programming, like teaching in other contexts, consists of two primary sets of activities: analysis and synthesis, with the former implying the partitioning of knowledge, information, and skills into components; and the latter implying the development and assembly of materials that will teach the content by building a sequence of questions or problems to move the learner toward the objectives (Stolurow, 1961.)

Efforts to generate a sound theoretical framework on the basis of past research in human learning are likely to yield minimal returns because most data from such research come from the study of groups of learners, not individual learners. Such studies have identified few interaction effects between learner variables and methods variables, yet it is precisely this kind of information that is most useful in making programs to meet the needs of the learner.

Recently, however, the notions of behavioral analysis, as they relate to human learning, suggest a promising approach having characteristics that may serve as a model. Lane has specified the features of the model by making the following points:

(a) The student's current behavior must be carefully assessed.

(b) The desired terminal behavior must be analyzed in detail.

(c) A program must be set down that will lead in small steps from initial to terminal behavior. (Lane, 1962, p. 115).

What are the desired terminal behaviors in visual space perception? Creating a program means taking on the task of specifying in precise behavioral language what the learner does (Green, 1962, p. 149). The subject area must be analyzed to determine the form of behavior desired at the end of the course. The material must then be structured as a large number of steps, beginning with assumed entering behaviors and built up gradually to the criterion behaviors. Each segment of content must evoke appropriate responses on the part of the student, and techniques must be worked out for prompting or otherwise encouraging the behavior at the appropriate time. As Hively (1962, p. 296) indicated, "learning is more efficient, in terms of either trials or errors, when training in a difficult discrimination is preceded by training in similar but easier discriminations, than when all training is given in the difficult discrimination alone." As the program progresses, the appropriate behavior must be brought under specific stimulus control. For an area as poorly defined as visual space perception this is admittedly a difficult task. The present state of the art of programming and its employment finds a considerable dependence upon error counts to estimate the degree of stimulus control desired.

Behavioral Specifications.

To ascertain the nature of the behaviors involved in a perceptual task such as performance on the Space Relations Test (DAT), a number of students were asked to verbalize their thoughts as they selected answers to each item. (The test is cast in multiple response form, meaning that for any given item from none to five alternatives may be correct.) These verbalizations were recorded on tape for later analysis. The analysis led to the designation of what seemed to be several discrete components essential to successful performance. Verbalizations by successful students showed the following features:

(a) rapid identification of the characteristic stimulus features of each alternative presented as a possible answer.

(b) identification of accurate space relations between observable and unobservable parts of the developed solids. (All sides of the solids could not be observed by looking at paper drawings.)

(c) ability to retain accurate orientation of the parts after the process of folding. In other words, successful examinees could somehow identify the proper space relations of the developed solid and could therefore identify those alternatives that failed to show such relations.

While several additional peripheral behaviors were detected in isolation, the above features seemed to constitute the core used by the successful performers and were most often lacking in the efforts of unsuccessful performers.

Program Content.

The large task that loomed at this point, of course, was to select means through which the desired behavior might be developed. Realizing that a considerable amount of student time would be needed to acquire the desired response patterns, it became convenient to work within the framework of an established course. It was decided that the required learning could be realized most logically through a short course in elementary geometry. The selection of content for the course was guided by Mathematics for Junior High School, a two-volume edition prepared under the supervision of the School Mathematics Study Group (1960). Among the topics included were the basic elements of point, set, line, line segment, ray, angle, simple plane figures (triangles, quadrilaterals, circles), and simple solids. It should be remembered that the usual approach to the teaching of geometry has not added significantly to performance on visual space perception tests (Ranucci, 1952; Brown, 1954). When one realizes, however, that the emphasis in the teaching of geometry is usually on development of formal proofs based on given information, the failure to add to the performance on spatial perception measures is not surprising. The behavioral demands are simply different.

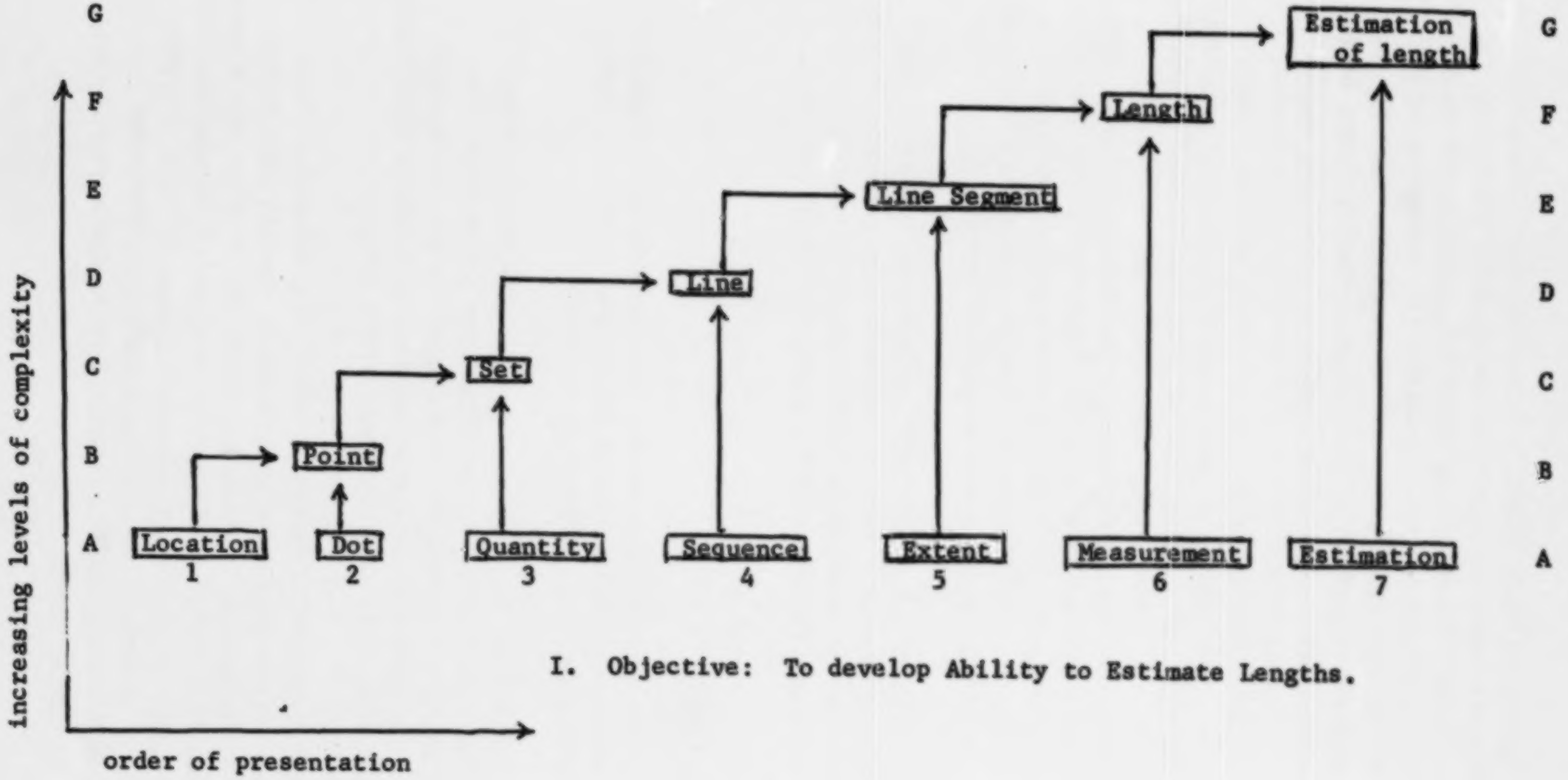
A course outline (see Appendix A-1) proceeding from the simple to the more complex behaviors defined the main body of content. In addition to the information provided by the analysis of taped protocols of successful and unsuccessful examinees, information from other sources was used (Shettel, *et al.*, 1961, p. 16).

Program Organization Chart.

The conceptual framework used as a guide for programming was developed into a series of charts similar to the "lattice structure" (Woolman, 1962.) A total of 15 charts were constructed. (See Appendix A-2).

Figure 1 shows the chart used to construct program frames aimed at helping students to improve their accuracy in estimating lengths of line segments. The major objective is specified in the upper-most cell on the right. The designation of a specific objective imposes a definite specification for every element entering into that segment of the program.

As with the "lattice structure" (Woolman, 1962), the sequential presentation of elements moves from left to right, columns one through seven. The baseline cells (Row A) represent information which the student was assumed to possess before starting the section. It may be designated as the required entering behavior. In Figure 1 this would include a grasp of such concepts as location, dot, quantity, sequence, extent, measurement, and estimation. The order of baseline elements is guided by the objective for a particular segment, represented on a single chart.


G
F
E
D
C
B
A
increasing levels of complexity
order of presentation
Location
1
Dot
2
Quantity
3
Sequence
4
Extent
5
Measurement
6
Estimation
7
Point
Set
Line
Line Segment
Length
Estimation of length

I. Objective: To develop Ability to Estimate Lengths.

Figure 1

A Chart to Schematize Relationships for Estimation of Length

Programming began with frames indicating location by using a dot to specify a particular point. The concept of point was related to the next baseline concept, quantity, and reference was made to a quantity of points or a set of points. This higher level concept when related to the next baseline concept, sequence, led to development of a set of points in a sequence which could represent a line. Combining the concept, line, with the next baseline concept, extent, the concept of line segment emerged as a line of a given extent. Then measurement was introduced. The measurement of a given line segment led to the length of that segment. Having a given line segment, the student was to judge (or estimate) whether a second line segment was longer than, or congruent with, the first line segment. Thus using concepts of length and estimation, the objective of estimation of length in comparison with a given standard was reached.

With the help of the chart the relations emerge in graphic form. The number of cells and levels may vary depending upon the purpose and level of complexity. The cells increase in complexity by slight increments and organize the body of the subject matter to expose graphically the areas of similarity and differences among overlapping concepts. While the chart specifies particular elements of content, it provides a great deal of flexibility in frame writing. The structure provides a convenient and explicit organization of the interrelated lower order terms, and furnishes a basis for generating higher order terms in a systematic way. The purpose of such a chart is best stated by Woolman:

> The major objective of the lattice approach is to insure that the student makes the proper lower order choices, and assigns proper emphasis when higher order terms are used. This is done through careful programming of the lower order terms so that the higher order term is perceived as a "resultant" of a given set of lower order relationships. A higher order term, under these conditions, becomes a label for the set of lower order relationships previously presented. (Woolman, 1962, p. 37)

While a program gradually becomes more complex as it progresses, later material is not necessarily more complex at every stage than earlier material. Complexity of terms and difficulty of frames need not correlate very highly. Once a higher level concept has been presented, it may then be used as a baseline cell in subsequent charts. The learner must, of course, continue to demonstrate control over the current material and relate it to all previous cells in the chart.

Frame-Writing and Testing.

Whether responses should be of the "constructed" type or "multiple-choice" type appeared to be a moot question. After considering arguments, pro and con, it was decided to follow the suggestion of Margulies (1962) and use the "constructed response" primarily in the early stages of learning while employing "multiple-choice" items where discriminations between stimulus conditions previously encountered were to be sharpened.

While specific procedures for program-writing become hazy in detail, several sets of general techniques have been offered. What may be referred to as the "classical" program offered by Holland (1962) and Skinner (1958) was probably most influential in the development of the present program. Application of some of the more important principles is treated in the following paragraphs.

The principle of immediate reinforcement leads to the stimulus-response-feedback sequence. To facilitate program development along these lines, index cards (size 3 x 5) with stimulus on one side and feedback or correct response on the back were used. The student could read the stimulus material, write his response on a separate answer pad, and turn the card over to check his response. A practical assist was made of a "flip-file" (a series of 25 plexiglass pockets for insertion of the index cards). This device aided in keeping the cards arranged in their proper sequence; it prevented the students from writing on the cards; and it permitted addition of frames, removal of defective frames, and rearrangement of cards without disturbing the remainder.

The gradual progression to more complex concepts was portrayed in the charts (see Appendix A). Considered in this progression was the behavior repertoire to be acquired and the density of stimulus support provided in successive steps. The objective was the emission of adequate responses (low error rate) in the presence of minimal cueing. In accordance with this principle, efforts were made to "wean" the student gradually from dependence upon maximum support in order to reach frames commensurate with the terminal conditions.

To establish a concept, extensive discrimination training was provided. Various examples were generated so that, having made a response in a variety of contexts, the learner was trained for retention of concepts. To extend opportunity for discrimination training, a series of panels in a separate booklet and an object kit of solids were provided. The panels consisted of drawings of various geometric designs as well as cut-out patterns of solids. The object kit was a box containing solids in the form of cubes, rectangular solids, pyramids, cones, and variations, or combinations of these. Both the pattern-folding and solid-object manipulation were intended to provide tactual-kinesthetic as well as visual feedback for the learner.

Preliminary testing helped to identify items that were ambiguous or too difficult, to locate gaps in the program, and to indicate where the program had moved too rapidly. Much attention was given to item testing in the early stages. After a series of 15 or 20 items were written, an adult (elementary teacher) was asked to go through the items to check for ambiguity and to give a general critical evaluation. After revision, the set of items were then tested by using students at the appropriate grade level. For most of the initial phases of development, students at the University School (The University of Michigan) donated time to participate in the testing phase. The availability of comprehensive test records made possible the selection of those students with low performance scores on earlier spatial measures. Before the student went through the program frames, he was encouraged to request clarification at any point where he was uncertain. The additional clarifications provided to the student were noted and incorporated into the revised format. As might be expected, the first subject through the program usually generated the most alterations. By the time the frames were presented to the fourth and fifth students, relatively few changes seemed necessary.

In the above manner, 493 frames were developed, as well as the panel booklet containing a series of 21 panels of figures and patterns, and the object kit of 11 solids. Up to this time, the program had been tested and validated with tenth-grade students. It was suspected that a younger student could probably deal effectively with the program, if he possessed a reading comprehension level equivalent to that of an average seventh grader.

To explore the adequacy of the program at the eighth-grade level, a sample of seven subjects was used during the summer of 1962. At the commencement of classes in September, 1962, these subjects were to be in the eighth grade in the Ann Arbor Public Schools. The Space Relations Test (DAT) was to be administered both before and after exposure to the program.

While the primary purpose of the exploration was a further evaluation of the difficulty level of the program and of the time required to complete it, and while the data did not justify any elaborate analysis, it was gratifying to find an average gain of slightly over one standard deviation for the experimental subjects.

Revised Program Design.

As a result of the addition of several review frames and the subdivision of a number of lengthy frames, the final version of the program consisted of a total of 505 frames. At this point it was decided that the program should be adequate for testing on a larger scale. To facilitate the presentation of the program to a larger group, the materials were organized into ten units varying from 50 to 53 frames each.

Each of the units was compiled into a separate booklet of 10 or 11 pages. The program was linear in design; it required both constructed and multiple-choice responses; and it employed the horizontal format with five levels per page. A separate answer sheet, lines numbered to correspond with frame num-

Table 1

Program Composition

Program Unit	Number of Frames	Instruction
1	50	Basic terminology and practice in estimation of lengths.
2	50	Additional terminology and differentiation between discriminative cues.
3	50	Discrimination between plane geometric figures.
4	50	Integration of parts into plane pattern.
5	50	Introduction to solids; development from parts in plane pattern.
6	50	Development of cubes and pyramids from patterns.
7	50	Development of cones and circular cylinders from patterns and practice in differentiation between classes.
8	53	Differentiation within a class; all cubes, but varying patterns--relative position of markings in pattern and folded form varied.
9	50	Practice in judging position of shadings in solid from relative position in pattern and effects of rotation.
10	52	Judging relative positions of various markings by projecting self within framework of solid.

bers, was provided with each booklet. Thus the student could read the initial frame, write his response on the answer sheet, turn the page to compare his response with the answer given, and in this manner proceed through the first level. After completing the top level, the student was directed to return to page one and work through the second level. This procedure continued until the fifth level was completed. Also provided for each student was a panel booklet and an object kit (each object was clearly numbered in a separate compartment in a box) to be used for reference when the program required it.

THE FILM STRIP

The purpose of this instructional component was to provide a series of pictures illustrating the development of flat patterns into solids. Initial frames depicted two boys in a store preparing to flatten two large boxes to make room for further unpacking. The boxes were identical cubes. But after disassembling the two solids the boys noted two unlike flat patterns. They wondered why different arrangements of the square surfaces as flat patterns could produce the same cubes. The problem was then set to show how the parts of a cube can be arranged in various patterns and still be developed into a cube. By picturing the various positions of the six squares that could be folded into cubes, the rules of development were identified by induction. Illustrations were then given showing that when the rules were obeyed the patterns could be developed into cubes and that when the rules were not followed it was impossible to produce the solid. While most of the frames dealt with cube patterns, other patterns which produced cones, pyramids, and spheres were also shown.

A taped sound track with appropriate comments was run in synchronization with the film strip. A total of 81 frames was made.

A unique feature of the filmstrip included silent cueing to change frames at 1/20th of a second without interrupting the flow of narration. Where desirable to show development as a nearly continuous motion, the effect of animation could be produced.

TACTUAL-KINESTHETIC MATERIALS

A kit of eleven solids-cones, pyramids, and rectangular solids placed in cubicles in a box-was provided for each pair of students during programmed learning. The program referred to the solids and required the student to manipulate them in various ways to clarify certain perspectives. In addition, a total of 21 panels was used in connection with the program. Thirteen flat patterns were inserted in pockets within the panel folder. The plates below show the solids used in the kit.

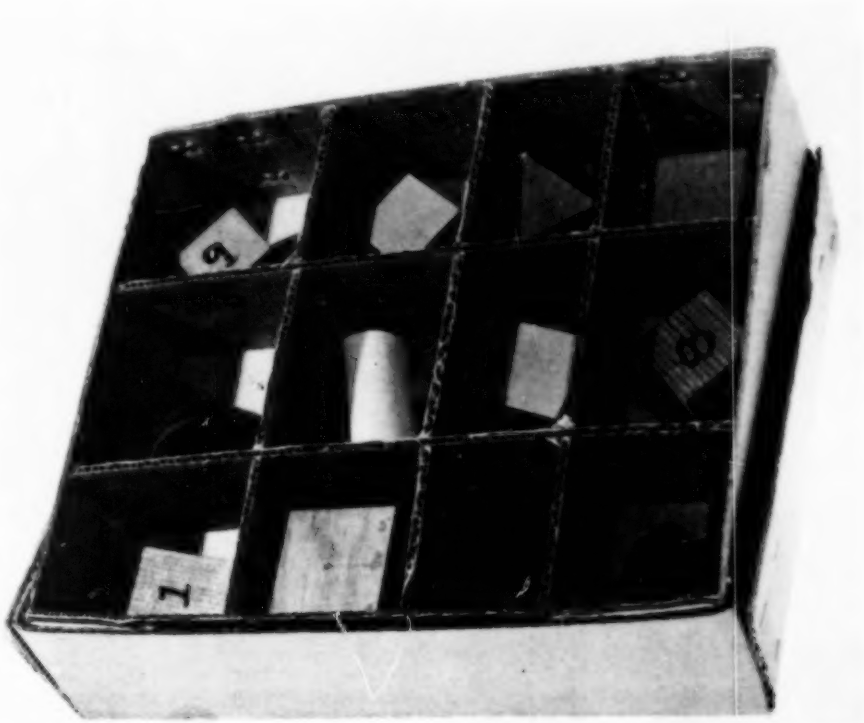


Plate I

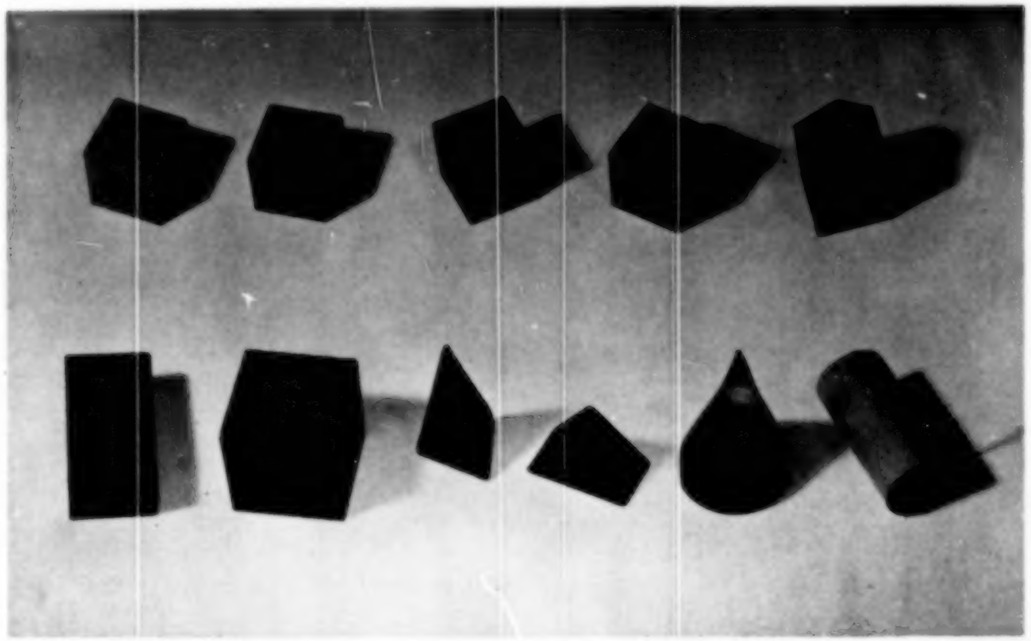


Plate II

The Solids in the Kit

Numbers and Corresponding Solids	Initial Reference in Program
1 - Rectangular	235
2 - Cube	259
3 - Triangular Pyramid	281
4 - Quadrangular Pyramid	289
5 - Circular Cone	310
6 - Circular Cylinder	318
7 - Cube, 2 adjacent faces shaded	352
8 - Cube, 2 opposite faces shaded	356
9 - Cube, 3 consecutive faces shaded	383
10 - Cube, 3 adjacent faces shaded	391
11 - "House," 2 opposite "roof" faces shaded, with 2 alternate, opposite square faces shaded.	472

Special information and problems were posed in the panels and where the student needed objects to observe and to manipulate appropriate reference was provided.

SAMPLING

Subjects for the experiment were drawn from schools in four communities in or near Greater Detroit. Total population of the communities ranged from about 2,000 people to over 76,000.

Research which must depend for subjects for its experiments upon the voluntary cooperation of schools is usually less than ideal in terms of sampling. In the present study, it was necessary to accept those classrooms offered by the school administrators. Hence, random sampling according to individual subjects could not be made. Selection was further limited by the fact that materials, especially the criterion tests, were suitable for students in the eighth grade and beyond. And since it was desirable to select classes in mathematics because of relevance between the programmed content and the subject matter, a still further limitation on sampling was imposed.

Because of the above restrictions, the sampling procedure did not provide groups that could be reasonably assumed equal at the beginning. Also, differences between schools were such that across-schools analysis could not be performed defensibly by pooling all subjects under the same treatment and proceeding to draw contrasts with other pooled groups under other treatments.* Consequently, the study was conceived as five separate experiments with different samples, meaning that results were analyzed within schools and not across schools.

Another fact that determined the type of analysis was that mathematics classes in the schools involved were differentiated on the basis of past achievement. In other words, it was found that students who aspired to higher education and who had shown adequate achievement in mathematics were typically found in "advanced" classes, while students who did not plan on going to college, and who had only mediocre achievement or below, commonly were found in classes labeled "general mathematics" or "practical mathematics." The above differences suggested that analysis of co-variance by treatments within schools was the most desirable procedure for determining significance of the data.

---

*This restriction applies particularly to analysis as originally intended which was changed to the analysis covariance so that differences in the samples could be adjusted for comparison.

Ypsilanti: East Junior High School

Ninety-four eighth-grade students were chosen from East Junior High School, Ypsilanti, Michigan. Four classes in mathematics participated in the pilot study to evaluate the adequacy of the program. Two teachers were involved, each instructing two classes. There were 49 boys and 45 girls. A control section had 18 boys and 18 girls, while the experimental groups had 31 boys and 27 girls.

East Junior High School was chosen because its students represent a good cross section of the community, which contains a fairly even distribution of solid-economic classes, from professional and managerial people at the "privileged" end to unskilled laborers at the other. Student population of the school was about 700

St. Clair Shores: Ottawa Junior High School

This community is within the boundaries of urban Detroit and has a population of about 75,000. The total number of students attending the junior high school was 783. Of that number, 203 participated in the experiment. All Ss were in the ninth grade. Twenty-eight Ss made up the control group and 175 received experimental treatments including the program, film, and film plus the program. Ss came from families that heavily represent the middle-middle socio-economic class.

Walled Lake Junior High School

Walled Lake, Michigan has a total population of about 3,500 people with a socio-economic range from household service workers to professional people and top-grade technicians with about an even distribution of economic classes between the two extremes.

Of the 918 students in this school, 585 were involved in the study. Both eighth and ninth grades were represented, 229 in the eighth grade and 256 in the ninth grade. Among the eighth graders, 48 were in the control group and the remainder were distributed among groups that received instruction. In the ninth grade, 40 Ss were in the control group and the remainder, 216, were in the various treatment sections. For purposes of analysis, Walled Lake furnished Ss for two separate experiments identified by grade, eighth and ninth.

Saline High School

Saline, Michigan is a community of slightly over 2,500 people. The school system is a consolidated one that serves an outlying rural area. The total number of students in the high school were 618, ninety-nine of whom participated in the study. Twenty-eight Ss composed the control group and 71 were divided among the indicated treatments.

THE YPSILANTI PILOT STUDY

The purpose of the pilot study was to examine the effectiveness of the program and to make alterations suggested by the results.

Ninety-four Ss at East Junior High School, Ypsilanti, Michigan participated. The school provided a sample of students from low- to middle-income families. There were four classes in eighth-grade mathematics with approximately 24 students per class. Two of the four classes were assigned by random selection to the experimental treatment, while the remaining two classes served as control sections. The experimental group received the instructional program in addition to the pre- and posttests; the control group received only the tests and not the program.

From the total N, 27 matched pairs were identified on the basis of equivalent scores on the Space Relations Test (Form A) of the Differential Aptitude Tests. Available scores on the Lorge-Thorndike Intelligence Tests and seventh-grade mathematics grades did not differentiate significantly between the experimental and control groups.

DATA COLLECTION PROCEDURES

To minimize a possible "Hawthorne effect" (French, 1953), efforts were made to collect data under conditions corresponding as closely as possible to the school's regular program. For the administration of both the pretest and posttest Space Relations Test, all students were directed to go to the school cafetorium (combination cafeteria and auditorium). Here the students were seated at tables, two students per table, at diagonally opposite corners. One of the school's counselors was in charge of the test administration and the teachers of the classes involved acted as proctors. This practice was the school's routine group-testing procedure. All materials and directions for administration were supplied by the investigator (see Appendix B and C).

The experimental procedure was restricted to the normal 47-minute class periods devoted to mathematics over a period of about three weeks. It was administered throughout by the teacher who was normally responsible for teaching the class.

Initially, the teachers in both the experimental and control group administered a pretest consisting of sample items from the instructional program, and referred to as the Geometric Inventory, to ascertain the background knowledge already possessed by the subjects. Each teacher had been advised not to discuss the content beyond the directions provided (see Appendix C). Hereafter, the teacher in the control class continued with the regular mathematics course. The teacher in the experimental group introduced the students to programmed instruction and explained the practices to be followed in utilizing the materials (see Appendix E). Each student was supplied with the first programmed unit together with an answer sheet. He was instructed to record on the answer sheet for each unit his starting and completion time, plus interruption and resumption time if only partially finished at the end of a class period. Students proceeded to work through the program individually without help and at their own pace. All remaining units, plus supplementary materials, were placed on a table at the front of the room. Here the student could select his lesson materials as they were needed. A few minutes before the end of each class period the teacher alerted the students to indicate the time on the answer sheet and to put their lesson materials into their individual folders in a filing cabinet. This procedure was already established, having been followed in their mathematics course. In this way, the student could resume where he had left off without difficulty; also, all of the materials remained within the classroom. After the students had completed the programmed course, the posttest Space Relations Test (Form B) was administered in the manner described earlier. Following completion of the principal criterion test, the posttest, the Geometric Inventory was administered by the classroom teacher during the next class session. The final data-gathering device, presented to the experimental group only, was a questionnaire designed to reflect the attitude of the individual toward the program. All materials were returned to the investigator for scoring and evaluation.

VARIABLES

Space Relations Test

The principal dependent variable was the individual's ability or skill in spatial visualization. Admittedly this ability consists of complex psychological components. While space tests may not be precise replications of tasks involved in object visualization, they nevertheless contain "tasks which are psychologically similar" (Thurstone, Thelma 1961).

The Space Relations Test of the Differential Aptitude Tests presents two-dimensional patterns with some of the sections of the patterns having special markings. Each pattern is accompanied by five sketches of three-dimensional figures. The examinee is to select those figures which could be made from the original pattern. At least one figure is correct and in some instances all five alternatives may be correct. The figures are related, requiring the examinee to imagine the movement of parts in the development of solids from the flat patterns. Forty such items are included within the test to be completed within the 30-minute time limit.

This test appears to require the subject to recognize the identity of an object when it is seen from different angles, to imagine the movement or internal displacement among the parts of the configuration, to perceive the spatial pattern accurately, and to remain unconfused by the varying orientations in which the pattern is presented. In addition, equivalent forms of the test are available (Form A and Form B), so that the same test need not be used for both pretest and posttest.

The reliability of the Space Relations, (Form A), is reported as 0.93 for eighth-grade boys and 0.86 for eighth-grade girls. The split-half coefficient, corrected by the Spearman-Brown formula, was used in the computation. According to the Test Manual (third edition), "the data for Form B are substantially the same as those for Form A; any judgments based on Form A data would hold for Form B" (Bennett, Seashore, and Wesman, 1959).

As a criterion, the Space Relations Test may be viewed logically as a test of the degree of transfer of learning resulting from the instructional program. No items from the test were included in the program; in fact, Form B, the posttest, was not viewed by the programmer until after the program was constructed.

Content-Learning Measure

In addition to the Space Relations Test, a special achievement test was constructed based only on the program content. It may, of course, be argued that performance on the program *per se* provides evidence of the effectiveness of learning. This may not indicate, however, the availability of responses when the stimulus items are not embedded in the sequential structure of the program itself. Thus a sample of criterion items taken directly from the program was assembled. In effect, this could be viewed as a terminal unit of review frames lacking immediate feedback and with low cueing density. A total of 60 such items calling for 66 responses were included in the sample and these were administered to the subjects in a four-page booklet referred to as the Geometric Inventory (see Appendix D.) Both multiple-choice and "self-constructed" responses were required by the items. Responses were entered on a separate answer sheet provided with the booklet. In view of the low scores on the pretest, a second form was not deemed necessary. The same form was therefore used as a pretest and posttest for both the experimental and control groups. A test-retest correlation of 0.80 for the control group was found. No time limits were imposed but the average completion time was approximately 30 minutes.

Additional Measures

To study possible significant effects of residual factors on performance on the instructional program and on the Space Relation Test, the Iowa Tests of Basic Skills, administered as a part of the school's regular testing program just prior to this study, provided a useful source of data. The Iowa Tests of Basic Skills differ from most other achievement test batteries in that they "are concerned only with generalized intellectual skills and abilities and do not provide separate measures of achievement in the content subjects" (Iowa, 1956, p. 10) One of the tests of possible significance to performance on a verbal program is a "Test R: Reading Comprehension." This test consists of selections which vary in length from a few sentences to a full page. An attempt was made to represent as completely as possible within this test every type of material encountered by the pupil in his daily reading. The reading process is here viewed not only by the extent to which the individual "apprehends the author's meaning, but also on the degree to which he grasps the significance of the ideas presented, evaluates them,and draws useful conclusions from them" (Iowa, 1956, p. 54).

A second test in this battery with potential relationship to the instructional program content is "Test A-1: Arithmetic Concepts." The emphasis on this test is on "understanding of the number system, of terms, processes, and operations, and of units of measurement" (Iowa, 1956, p. 69). Among the concepts included are the following:

> Concepts Involving Standard Measures: telling time and time zones; knowledge of calendar; reading thermometer; use of ruler; changing from one unit to another; performing fundamental operations with numbers; estimating quantities.
>
> Concepts Involving Geometric Figures: recognizing kinds of geometric figures; perimeters and areas of plane figures; parts of a circle; angles and triangles; use of protractor and compass" (Iowa, 1956, p. 70)

The reliability, at the eighth-grade level, for the "Reading Comprehension Test" is given as 0.91 and for the "Arithmetic Concepts Test" as 0.85. The coefficients were computed by the split-half procedure and were based on 500 cases drawn at random from the complete standardization sample at that grade level.

The Lorge Thorndike Intelligence Tests had been administered to all students in the school system. This test provided both verbal and nonverbal scores. The Verbal Series consists of subtests, each using the verbal medium, labeled as follows: "Word Knowledge," "Sentence Completion," "Verbal Classification," "Verbal Analogies," and "Arithmetic Reasoning." The authors of the test suggest that such tasks "provide a good and useful measure of ability to deal with abstractions presented in verbal form, and that, since this type of ability is related to academic achievement, they provide a good index of scholastic aptitude" (Lorge and Thorndike, 1955).

The Nonverbal Series is made up of "Figure Analogies," "Figure Classification," and "Number Series." It is indicated that for the average child these tests will not predict school performance as well as scores based upon the Verbal Series but that they should "give significant predictions of aptitude for some types of work with requirements for spatial thinking" (Lorge and Thorndike, 1955, p. 1). Use of both series is recommended for the appraisal of children in schools. For this reason the combination, or total, was considered to be the most satisfactory score to use for comparing and for constrasting groups.

Although the tests are given with definite time limits, the Lorge-Thorndike battery are primarily power tests. The alternate forms reliability of these tests shows the Verbal at 0.90 and the Nonverbal at 0.81. A total of 724 cases were used in these computations.

To sample the reactions of the subjects toward the programmed instructional materials, a short questionnaire was devised. Statements of opinion about programmed instruction were given to which the students indicated their degree of endorsement on a Likert-type scale typically running from "strongly agree" to "strongly disagree." Other comments on learning through use of the program were encouraged.

SUMMARY OF PILOT PROCEDURES

The sample of Ss was provied by four eighth-grade classes of approximately 24 students each (total N was 94) from a public school in Michigan serving low- to middle-income famileis. The classes were assigned randomly to the experimental and control treatments with the former group receiving the instructional program in addition to the pre- and posttests. The control group received only the tests and not the program.

For purposes of analysis in this pilot study, 27 matched pairs were selected on the basis of the pretest on the Space Relations Test of the Differential Aptitude Tests. Matched groups did not differ significantly on pretests involving: the Geometric Inventory, an achievement test constructed to measure learning by use of the program, scores on the Lorge-Thorndike Intelligence Tests, and mathematics grades received in the seventh grade.

Data collection was conducted under conditions corresponding as closely as possible with the normal activities of the school. Classroom management of the programmed materials allowed for self-pacing.

PROCEDURE FOR THE EXPERIMENT ITSELF

The same procedure was followed as described in the pilot study except for the following:

1. Minor editorial changes were made in the program. The ten units were bound into one book of reduced size.

2. Instructional components were added as indicated below:

a. For the St. Clair Shores samples the motion picture was added to test its motivational effect.

b. For the Walled Lake and Saline Samples all instructional components were used in the experimental sections. The full complement of components included the program, film, and filmstrip. (The tactual-kinesthetic materials were always used in conjunction with the program.)

3. No matching procedure was used because it was decided to use the analysis of covariance to handle whatever initial differences might be found between groups.

4. The entire 94 Ss, in the pilot study were used in the final analysis instead of only the 27 matched pairs. Thus, the Ypsilanti subjects supplied data for both types of analysis described in the "Results" chapters.

ST. CLAIR SHORES

The purpose of the St. Clair Shores experiment was to compare three experimental treatments with a control situation as shown in Table 2. According to the hypotheses and the rationale pertaining to them, it was expected that as instructional inputs increased, the correlations between pre- and posttests would decrease and that the adjusted means on posttests would increase.

TABLE 2

TREATMENT BY GROUPS FOR ST. CLAIR SHORES SAMPLES

GROUP	N	PRETESTS	FILM	PROGRAM	POSTTESTS
I	28	X			X
II	28	X	X		X
III	89	X		X	X
IV	58	X	X	X	X
TOTAL	203				

WALLED LAKE

All instructional components were used singly and in combinations as shown in Table 6.

One control group plus seven experimental groups were used in the eighth grade with a total N of 229. In the ninth grade, a total of 256 Ss made up one control group plus five experimental groups as shown in Tables 3 and 4.

TABLE 3

TREATMENT BY GROUPS OF WALLED LAKE SAMPLES, EIGHTH GRADE

GROUP	N	PRETESTS	FILM	PROGRAM	FILMSTRIP	POSTTESTS
I	48	X				X
II	29	X		X		X
III	41	X	X			X
IV	22	X	X	X		X
V	20	X			X	X
VI	18	X		X	X	X
VII	27	X	X		X	X
VIII	24	X	X	X	X	X
Total	229					

TABLE 4

TREATMENT BY GROUPS OF WALLED LAKE SAMPLES, NINTH GRADE

GROUP	N	PRETESTS	FILM	PROGRAM	FILMSTRIP	POSTTESTS
I	40	X				X
II	41	X		X		X
III	39	X	X			X
IV	50	X	X	X		X
V	49	X	X		X	X
VI	37	X	X	X	X	X
Total	256					

SALINE

Ss in the Saline High School were divided into five groups, one control group and four experimental groups with treatments as indicated in Table 5.

TABLE 5

TREATMENT BY GROUPS FOR SALINE SAMPLES

GROUP	N	PRETESTS	FILM	PROGRAM	FILMSTRIP	POSTTESTS
I	28	X				X
II	15	X	X			X
III	19	X	X	X		X
IV	14	X			X	X
V	23	X	X	X	X	X
Total	99					

Table 6 shows the experimental treatments by school, grade, and sex. An original N of 1097 cases was reduced to 881 because of inappropriate procedures taken by some teachers of experimental sections. Unfortunately this reduced the N in several columns particularly in those labeled "Filmstrip," "Film and Filmstrip," and "Program and Filmstrip." The press of time would not allow additional data gathering, which would have required several months delay because of negotiations needed to adjust the curriculum schedule to admit the experimental treatments. But since no attrition was invoked by analytic requirements all data not contaminated by errors in instructional management were used.

SUMMARY

A total of 881 seventh- and eighth-grade students in four Michigan schools located in or near Greater Detroit participated in the study. The purpose of the research was twofold: (a) to determine the amount of change in measures of reliability of a standardized test of space relations as a result of varying amounts of instructional input that intervened between pre- and posttests, and (b) to compare the effects of instruction in space relations by using posttest mean scores as the criterion. A second criterion test was especially constructed to estimate the amount of learning that could be attributed to the instructional treatments.

TABLE 6

EXPERIMENTAL DESIGN SHOWING TREATMENTS BY SCHOOL, GRADE, AND SEX

School	Control		Program		Film		Filmstrip		Film Filmstrip		Program Filmstrip		Program Film		Program Film Filmstrip	
	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.	Male	Fem.
Ypsilanti 8th Grade	18	18	31	27												
St. Clair Shores - 8	12	16	52	37		28							35	23		
Walled Lake 8th Grade	17	31	13	16	27	14	9	11	6	21	6	12	15	7	13	11
Walled Lake 9th Grade	26	14	26	15	20	19			25	24			28	22	15	22
Saline 8th Grade	11	17			10	5	10	4					14	5	10	13
Totals	84	96	122	95	57	66	19	15	31	45	6	12	92	57	38	46

Aggregate N = 881

A pilot study on 94 subjects (eighth-grade) tested the efficacy of a program for teaching students how to handle problems of development, that is, the making of solids from flat patterns. In conjuction with the program a tactual-kinesthetic kit of materials was used to help Ss to move from concrete manipulation of objects to mental manipulation in the solution of problems of development.

Following the pilot study, samples from three schools involving eighth- and ninth-grade students were selected. Because it could not reasonably be assumed that the school samples represented the same population, the design was arranged to regard each school as sufficiently unique to represent parallel experiments.

Before the first experiment, and after the pilot study, a 15-minute motion picture was produced which would positively influence student motivation. With a total N of 203 Ss in a St. Clair Shores school (urban Detroit), four groups were selected and treated differentially. Treatments were as follows: control (only pre- and posttests), program only, film only, film and program. The experimental design was set up to allow for analysis by the covariance technique.

At Walled Lake, Michigan, a total of 229 eighth-grade Ss were divided into eight groups, which provided a sufficient number to test the single and combined effects of three instructional inputs: program, film, and filmstrip. In the same school 256 ninth-grade Ss comprised six groups which allowed all treatment comparisons except two, namely, filmstrip and the combination of filmstrip and program.

At Saline, Michigan, 99 Ss were divided into five groups which received the following treatments: control, film only, film and program, filmstrip, and the combination of film, filmstrip, and program.

The filmstrip was designed to present both aural and visual stimuli demonstrating the development of various solids, with test problems and followed by presentation of correct solutions.

CHAPTER THREE: RESULTS OF THE PILOT STUDY

As indicated earlier, the pilot investigation sought information on two primary questions: did the program succeed in conditioning the specific terminal behaviors (content learning); and, did the program have an effect on the larger repertory of discriminative behaviors (visual space relations) of which the specific terminal behaviors were a subset?

PROGRAM PERFORMANCE

The difficulty level of a program may be an important factor in its effectiveness as an instructional medium. One indicator of the program's difficulty may be the speed with which a unit is completed.

Time Data

How much time did the Ss need for each of the units of the program? The units were roughly equal in length but reference to panels and objects probably influenced the rate. The nature of the data suggested that the median might be a realistic representation of the time used because the time measures clustered around a point with a few extremes, as indicated by the range. Table 7 summarizes the time measures. The median time represents only that spent on the program. It excludes time required to distribute materials and to administer the tests. The range was from 210 to 454 minutes. There were 505 frames in the 10 units of the program. The median speed was 1.5 frames per minute with a range of from 1.1 to 2.4 frames per minute. For the individual units the median speed ranged from 1.7 frames per minute on units VI, VII, and VIII to 1.25 frames per minute on unit III.

Error Rate

Table 8 summarizes the students' error rates. In addition to being a measure of the program's difficulty, the error rate may also be one of the criteria by which the quality of the program may be evaluated. With the distribution of errors for individual learners very definitely skewed, the median was selected as representative of the error rate. For purposes of error tabulation the frames were given equal weight; that is, although several frames called for more than one response, each part was considered a fraction of the total frame. The only exception was Unit II, in which several review items were placed a single frame but were considered separate entities. For all 10 units combined the error range varied from 8 to 117 per individual with a median of 42.5, or 8.3 per cent missed of a possible 510 errors for the entire

TABLE 7

TIME DATA FOR PROGRAM, UNITS I-X, FOR 28 EIGHTH-GRADE STUDENTS

		Actual Time		Frames Per Minute	
Unit	Frames Per Unit	Range (Min.)	Median (Min.)	Range	Median
I	50	26-59	35	1.9-0.85	1.4
II	50	26-56	35	1.9-0.89	1.4
III	50	20-57	40	2.5-0.88	1.25
IV	50	18-48	33	2.8-1.04	1.5
V	50	11-49	34	4.5-1.02	1.5
VI	50	13-46	29	3.8-1.09	1.7
VII	50	16-56	30	3.1-0.89	1.7
VIII	53	15-40	29	3.3-1.25	1.7
IX	50	19-58	31	2.6-0.86	1.6
X	52	15-61	33	3.5-0.85	1.6
All 10 Units (28 Ss)	505	210-454	333	2.4-1.11	1.5

TABLE 8

ERROR DATA FOR PROGRAM, UNITS I-X, FOR 28 EIGHTH-GRADE STUDENTS

		Number of Errors			Error Rate (% of Total Possible)	
Unit	Response Units[a]	Range	Mean	Median	Mean	Median
I	50	0-12	4.25	3.5	8.5	7.0
II	55	0-16	4.57	4.0	8.3	7.3
III	50	0-10	3.68	3.5	7.4	7.0
IV	50	0-11	3.71	3.0	7.4	6.0
V	50	0-18	5.78	4.5	11.6	9.0
VI	50	0-18	5.32	5.0	10.6	10.0
VII	50	0-16	5.75	4.5	11.5	9.0
VIII	53	0-20	4.89	5.0	9.2	9.4
IX	50	0-17	4.74	4.0	9.5	8.0
X	52	1-17	6.11	5.5	11.8	10.6
All 10 Units (28 Ss)	510	8-117	48.07	42.5	9.42	8.3

[a] Each frame was considered a response unit except in Unit II where several review items, placed into a single frame, were considered as separate entities.

program. On separate units the error rate ranged from a low of 7 per cent on Units I and III to a high of 10.6 per cent on Unit X, indicating a relatively consistent error rate.

Geometric Inventory Analysis

As a measure of content learning, the Geometric Inventory showed the releative gains of the two groups. Table 8 summarizes the performance of the two groups. The range suggests the degree of equality of the two groups on the pretest. On the posttest the lowest score of the experimental group was almost as high as the best score for the experimental group on the pretest. For the control group the scores remained relatively stable. A comparison of the two groups, shown in Table 9, showed no significant difference between the two groups on the pretest although the control mean was slightly higher. A comparison of the posttest means, Table 11, showed the experimental group mean more than double the control group mean. The difference was significant at the .001 level of probability.

FINDINGS PERTAINING TO HYPOTHESES

These analyses assumed that the eighth graders in the sample were independently drawn from a normal population in so far as their visual space relations ability was concerned. While the distribution of scores on the pretest was slightly skewed, the departure from normality was small. Studies indicate that some departure from normality does not invalidate the use of parametric methods (Walker and Lev, 1953, p. 143).

Hypothesis 1

To test hypothesis 1, a Pearson product-moment correlation was computed for pretest and posttest scores (raw score data) for the control group and for the experimental group separately. Results are summarized in Table 11. The pretest-posttest correlation of 0.91 (rho also was 0.91) for the control group is in line with average reliability coefficients for the Space Relations Test. For the experimental group the pretest-posttest correlation was 0.79 (rho = 0.78). The problem of obtaining a test of significance of difference between two correlation coefficients is complicated by the possibility that a relationship may exist between the correlations. Thus an approximation of the standard error of the difference was obtained by using a formula designed to consider this relationship (McNemar, 1949, p. 125). The difference between these correlation coefficients was significant at the .05 level (C.R. = 2.00).

TABLE 9

PRETEST AND POSTTEST RANGE, MEAN, AND STANDARD DEVIATION OF SCORES ON THE GEOMETRIC INVENTORY

Group	N[a]	No. of Responses Possible	Pretest			Posttest		
			Range	Mean	σ	Range	Mean	σ
Exp.	25	65	6-30	19.56	6.05	31-63	50.84	8.83
Control	25	65	6-31	21.52	7.08	3-39	24.32	7.82

[a] Two absentees in control group reduced pairs to 25.

TABLE 10

COMPARISON OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE GEOMETRIC INVENTORY PRETEST

Group	N	Mean	σ	σ_M	σ_{D_M}[a]	t	df	Level Sig.
Exp.	25	19.56	6.05	1.23				
Control	25	21.52	7.08	1.44	1.35	0.16	24	n.s.

[a] Standard error of difference for matched pairs.

TABLE 11

COMPARISON OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE GEOMETRIC INVENTORY POSTTEST

Group	N	Mean	σ	σ_M	σ_{D_M}[a]	t	df	Level Sig.
Exp.	25	50.84	8.83	1.80	1.76	15.07	24	.001
Control	25	24.32	7.82	1.60				

[a] Standard error of difference for matched pairs.

Hypothesis 2

The second hypothesis implied that the mean score of the experimental group on the posttest Space Relations Test would be significantly greater (<.05 level of significance) that the mean score of the control group. Table 12 summarizes the data of the precourse measures on the basis of which matching was accomplished. What slight differences were uncovered were in favor of the control group. On the Lorge-Thorndike Intelligence Test (Total) the control group mean was 107.21 against an experimental group mean of 106.38. Three pairs of students were not included in this comparison because scores for this particular measure were not available. On the Space Relations Test the mean scores were almost identical with a mean score of 36.07 for the control group and a mean score of 36.04 for the experimental group. In neither case do the differences approach significance.

When the equivalence of the two groups on the basis of intelligence and space relations scores was thus established, attention was directed to the posttest Space Relations Test scores. Table 14 presents the tabulated group data on posttest Space Relations scores. While the experimental and control groups had almost identical pretest scores, their posttest scores were quite different. The experimental group mean was now 54.22 compared with a control group mean of 39.26. This difference of almost 15 raw score points was significant at the .001 level.

To examine more closely the degree of improvement from pretest to posttest for the respective groups, raw score means were converted into percentile ranks based on the norms for the standardized test to provide a meaningful basis for comparing scores on the two test forms. The norms vary for both test form and sex. Table 15 shows that all groups experienced improvement in terms of percentile ranks in posttest performance over their pretest scores. Form B means are lower, thus accounting for the gain in percentile rank even though raw scores did not change appreciably as with the control groups. The experimental girls gained 30 percentile units (65th to 95th); the control girls gained 15 percentile units (65th to 80th). The boys showed a similar pattern: the experimental groups gained 25 percentile units (55th to 80th); control boys moved from the 55th to the 70th percentile although the mean raw score gain was less than three points.

ADDITIONAL FINDINGS

Differential Effects

Several additional findings merit attention. The first applies to the change in performance for students at various levels of entering behavior as indicated by initial scores on the visual space perception measure. To examine the varying effects which an instructional program may have on learners at

TABLE 12

CORRELATION OF PRETEST AND POSTTEST SPACE RELATIONS SCORES

Group	N	r	z	σ_{D_z}[a]	$z_1 - z_2$	C.R.	Level Sig.
Exp.	27	0.79	1.07				
Control	27	0.91	1.53	0.23	0.46	2.00	.05

[a] Standard error estimate for related correlation coefficients.

TABLE 13

EQUIVALENCE OF GROUPS ON PRECOURSE MEASURES USED AS BASIS FOR MATCHING PAIRS

Measure	Group	Mean	σ	σ_M	σ_{D_M}	t	df	Level Sig.
Intelligence	Exp.	106.38	10.72	2.23				
	Control	107.21	10.68	2.23	1.18	0.70	23[a]	n.s.
Space Relations	Exp.	36.04	21.21	4.16				
	Control	36.07	21.81	4.28	0.72	0.04	26	n.s.

[a] IQ measures not available for 3 Ss.

TABLE 14

A COMPARISON OF THE EXPERIMENTAL AND CONTROL GROUPS ON POSTTEST SPACE RELATIONS SCORES

Group	Mean	σ	σ_M	σ_{D_M}	t	df	Level Sig.
Exp.	54.22	22.87	4.49				
Control	39.26	25.45	4.99	3.58	4.18	26	.001

TABLE 15

GAINS FROM PRETEST TO POSTTEST AS INDICATED BY PERCENTILE NORMS FOR THE RESPECTIVE FORMS AND SEX

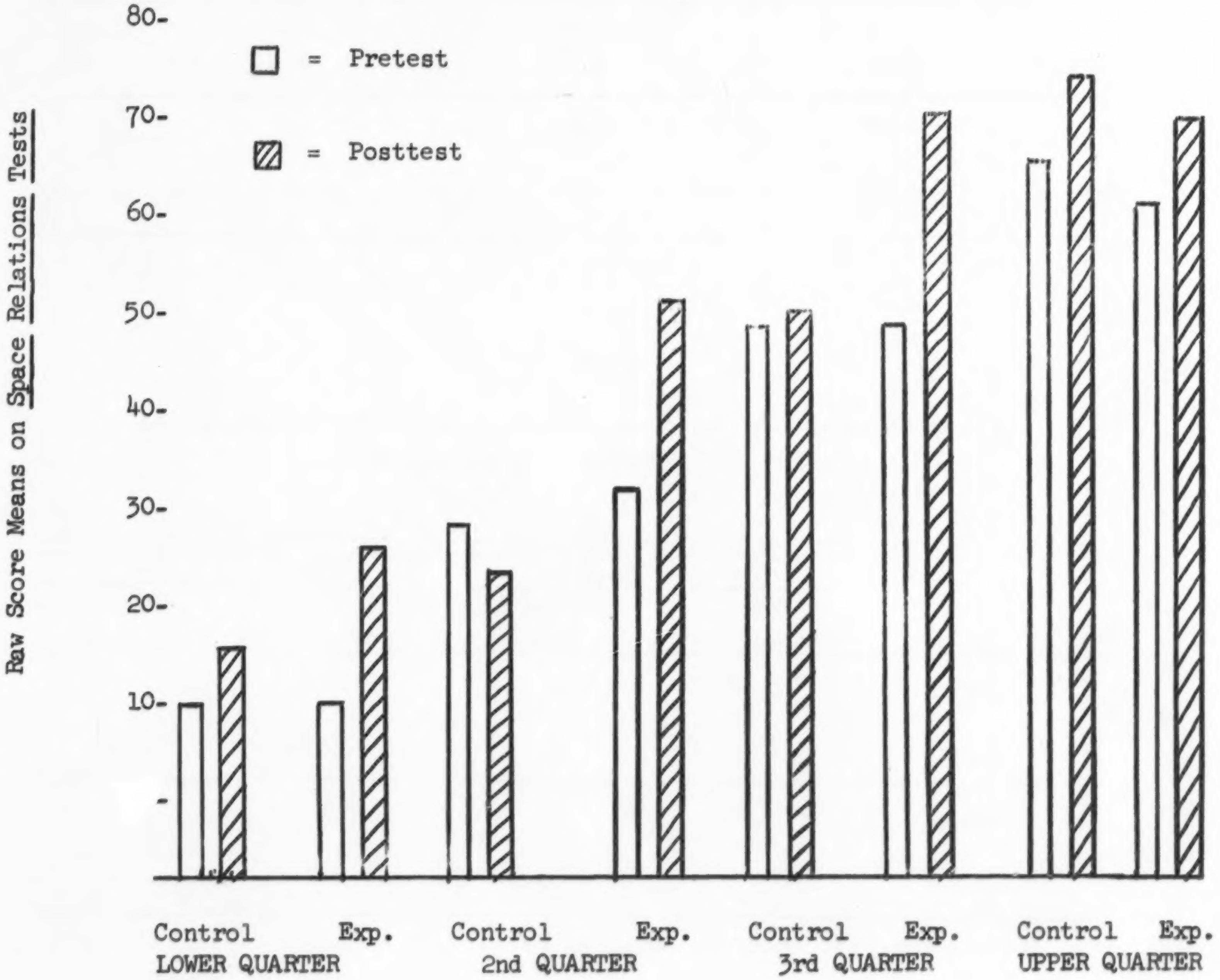
		Pretest (Form A)		Posttest (Form B)	
Sex	Group	Mean (Raw Score)	Percentile Rank*	(Raw Score)	Percentile Rank
F	Exp.	35.38	65	55.31	95
	Control	36.08	65	39.85	80
M	Exp.	36.64	55	53.21	80
	Control	36.07	55	38.71	70

*Percentile rank based on national norms.

several points on the initial performance continum, the sample was divided into quarters on the basis of pretest Space Relations Test scores; the resulting mean scores for each quarter, both pretest and posttest, are represented graphically in Figure 2. The graph indicates that for the upper quarter the experimental and control groups showed almost identical gains. Both groups in the lower quarter showed improvement, with the experimental group showing a slightly higher gain. Both groups may, of course, be viewed as regressing toward the mean. For the middle half (second and third quarters) the control subjects show the relative stability from pretest to posttest implied in the hypotheses tested earlier. In contrast to the stability of the

FIGURE 2

PRETEST-POSTTEST CHANGE FOR EACH QUARTER OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE SPACE RELATIONS TESTS



control group performance, the middle half of the experimental group shows marked gains from pretest to posttest.

Considering the possibility that the experimental treatment was proving most effective with the middle half of the sample, further evaluations of this group were undertaken. Hypothesis 1 was re-evaluated using only the middle half of the subjects in the computations. Results are shown in Table 16. The pretest-posttest correlation of 0.85 (rho = 0.84) for the control group compares favorably with the 0.91 correlation found earlier. However, the pretest-posttest correlation of 0.45 for the experimental group, when contrasted with the correlation of 0.79 for the entire experimental group, reflects a sizable drop in stability. The difference between the pretest-posttest correlations of these two sub-groups composed of the middle half of each group was significant at the .02 level (McNemar formula for standard error of difference was used).

TABLE 16

CORRELATION BETWEEN PRETEST AND POSTTEST SPACE RELATIONS SCORES FOR MIDDLE HALF OF SAMPLE

Group	N	r	z	σd_z	$z_1 - z_2$	C.R.	Level Sig.
Exp.	14	0.45	0.48				
Control	14	0.85	1.26	0.32	0.78	2.44	0.02

To examine the change in scores for learners at different levels of entering behavior, as indicated by pre-course scores on a measure designed to sample the instructional program content, the sample was divided into quarters on the basis of scores on the Geometric Inventory; the resulting mean scores for each quarter, both pretest and posttest, were represented graphically. As indicated in Figure 3, each of the four groups gained approximately 31 raw score points between pretest and posttest.

Relation to Residual Variables

Relationships of such residual factors as intelligence, reading comprehension, and arithmetic concepts to several of the variables studied in this research are indicated by the correlation coefficients in Table 17. Intelligence,

FIGURE 3

PRETEST-POSTTEST CHANGE FOR EACH QUARTER OF THE EXPERIMENTAL GROUP ON THE GEOMETRIC INVENTORY

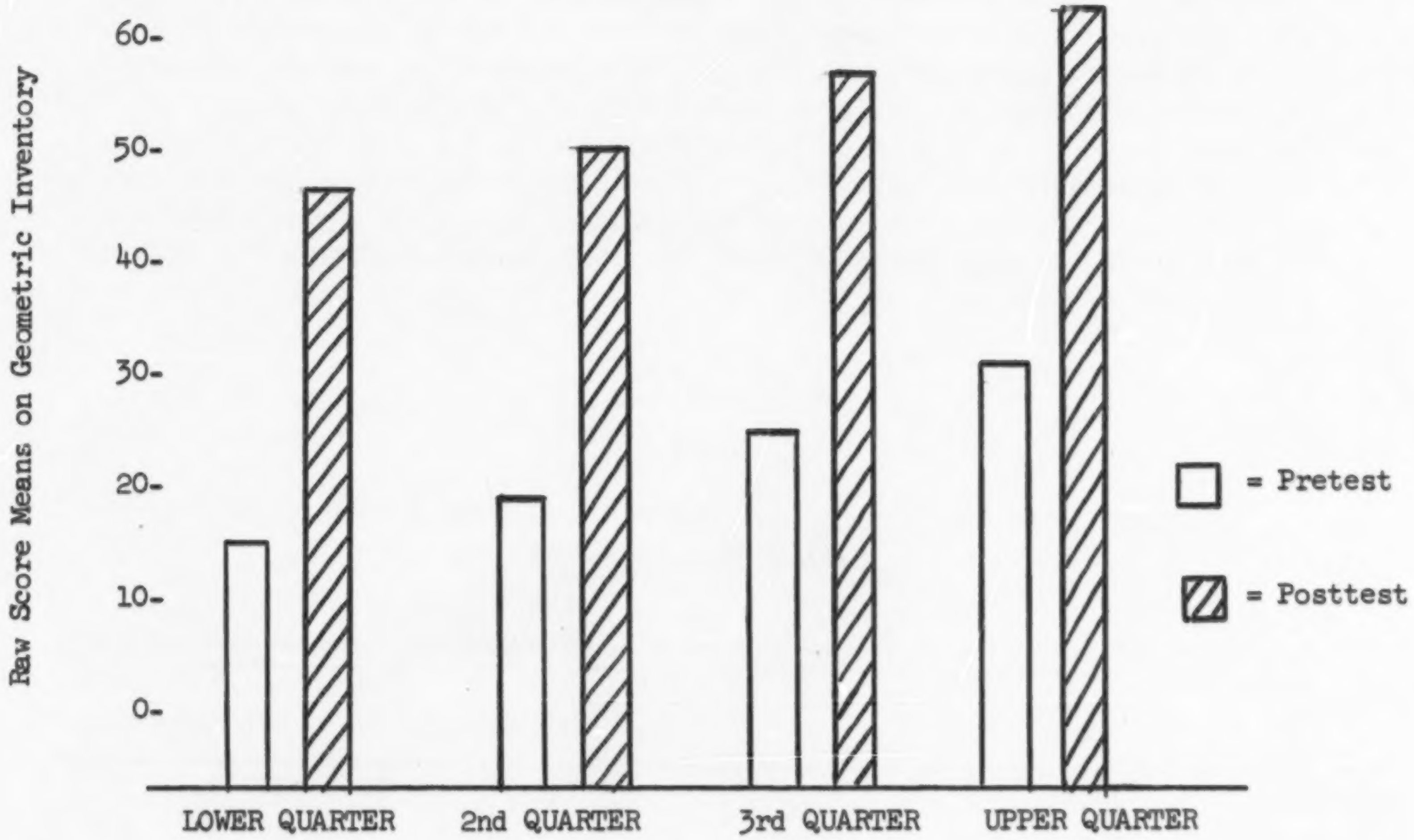


TABLE 17

CORRELATION OF PRE-EXPERIMENTAL MEASURES OF INTELLIGENCE, READING COMPREHENSION, AND ARITHMETIC CONCEPTS WITH EXPERIMENTALLY MANIPULATED VARIABLES FOR THE EXPERIMENTAL GROUP

	r	N	t	df	Level Sig.
LORGE-THORNDIKE INTELLIGENCE (Total) and:					
Pretest Space Relations	0.68	25	4.47	23	0.001
Posttest Space Relations	0.73	25	5.15	23	0.001
Pretest Geometric Inventory	0.72	25	5.00	23	0.001
Posttest Geometric Inventory	0.73	25	5.15	23	0.001
Error Rate on Program	-0.24	24	1.17	22	n.s.
Time on Program	0.18	24	0.84	22	n.s.
READING COMPREHENSION and:					
Pretest Geometric Inventory	0.55	27	3.30	25	0.01
Posttest Geometric Inventory	0.68	27	4.66	25	0.001
Posttest Space Relations	0.50	27	2.85	25	0.01
Pretest Space Relations	0.23	27	1.20	25	n.s.
ARITHMETIC CONCEPTS and:					
Pretest Geometric Inventory	0.44	27	2.45	25	0.05
Posttest Geometric Inventory	0.58	27	3.60	25	0.01
Posttest Space Relations	0.58	27	3.60	25	0.01
Pretest Space Relations	0.53	27	3.10	25	0.01

as measured by the Lorge-Thorndike Intelligence Test (Total) was significantly correlated with pretest and posttest measures of both the Space Relations Test and the Geometric Inventory. There was a modest, though nonsignificant, negative correlation between intelligence and error rate on the program. The time required to complete the program was not significantly related to intelligence, or to any of the other variables examined in this study.

The "Reading Comprehension" subtest of the Iowa Tests of Basic Skills was significantly correlated with the pretest and posttest Geometric Inventory as well as with the posttest Space Relations Test; however, there was no significant relationship with the pretest Space Relations Test.

The "Arithmetic Concepts" subtest of the Iowa Tests of Basic Skills was significantly related to both the pretest and posttest of both the Geometric Inventory and the Space Relations Test. For each measure the relationship was slightly higher on the posttest measures than on the pretest measures. Correlation coefficients are shown in Table 17.

Attitude Questionnaire

A questionnaire was administered after the instructional program and before any posttest results were revealed to check the possibility of a relationship between the effectiveness of this particular technique of learning and the students' attitude. The attitude survey revealed two statements receiving diverse reactions: Those Ss who expressed agreement with the statement, "Teachers can teach much better than a program," had more consistently scored below the median on the posttest Geometric Inventory; those expressing uncertainty or disagreement more often scored above the median. The obtained chi-square (two-tailed) value for one degree of freedom was significant at the .025 level (.02 as computed by Fisher's exact probability test) is indicated in Table 18. The second question receiving diverse reactions was, "How often would you like to use programmed learning?" Of those students desiring to use this medium over an extended period, the majority had scored above the median on the posttest Geometric Inventory, as shown in Table 19. On the other hand, the majority of those Ss indicating a preference for only occasional use or no use of programmed learning scored below the median. A chi-square of 3.88 was significant at the .05 level (two-tailed) for one degree of freedom. The same significance level was attained when using Fisher's exact probability test (Siegel, 1956).

Responses to the remaining items on the attitude survey were generally in agreement. Most of the Ss agreed that programmed learning was a good way to learn, that it was challenging, and that they could learn a great deal by using a program. Nearly all students disagreed with the statements that they did not have to think when learning with a program or that programmed learning was

TABLE 18

RELATIONSHIP BETWEEN PERFORMANCE ON THE POSTTEST GEOMETRIC INVENTORY AND EXPRESSED ATTITUDE TO STATEMENT, "TEACHERS CAN TEACH MUCH BETTER THAN A PROGRAM"

Score on Posttest Geometric Inventory	Attitude Agree	 Uncertain or Disagree	Total
Above the Median	2	12	14
Below the Median	9	5	14
Total	11	17	28

Chi square = 5.40 (continuity correction used); sig. at .025 for 1 df.

TABLE 19

RELATIONSHIP BETWEEN PERFORMANCE ON THE POSTTEST GEOMETRIC INVENTORY AND EXPRESSED DESIRE FOR DEGREE OF USE OF PROGRAMMED INSTRUCTION AS AN INSTRUCTIONAL INPUT

Score on Posttest Geometric Inventory	Degree of Use Extended	 Little or None	Total
Above Median	8	6	14
Below Median	2	12	14
Total	10	18	28

Chi square = 3.88 (continuity correction used); sig. at .05 for 1 df.

a boring method of learning. It was interesting to note that not a single student considered the program too difficult, and at the same time not a single student thought the whole program too easy.

Many of the findings presented in this chapter stimulate considerable interest per se, but speculation still remains on their meaning and the attempt to relate that meaning to currently accepted theory. Furthermore, empirical results usually have implications for future research and for practical application.

CHAPTER FOUR: COMBINED RESULTS

Data pertaining to the first hypothesis will be presented for all groups followed by data relevant to the second hypothesis.

THE FIRST HYPOTHESIS

The initial prediction was that reliability coefficients of the Space Relations Test would be comparable for all groups. The null hypothesis predicted that differences between pairs of correlations, representing groups under different treatments, would not exceed those attributed to chance. The hypothesis was posed with the expectation that results would justify its rejection on grounds given earlier.

Tables 20 and 21 list the correlations between pre- and posttest scores, Space Relations Test, according to treatment and sex within each grade and school. The range of correlations for all control groups overlaps considerably with the range of correlations for groups who received programmed instruction. Because of such overlap and because of the low N's, the number of significantly different correlations were found to be negligible.

Upon closer inspection, however, an apparent trend led to the pooling of groups by treatment, by school, and by sex. The pooling increased N's and provided a means for examining the significance of the trend detected by inspection of Tables 20 and 21. Table 22 presents pooled data by treatment, school, and sex for both the Space Relations Test (DAT) and the Geometric Inventory (GI).

Table 23 shows the calculated differences between correlations, using the z-transformation, by pooled aggregates: treatment, school, and sex. Eight of the possible 28 differences were significant at the .05 level of probability. One of those eight significant differences was counter to prediction, that is, groups which experienced all inputs (FFSP) had a higher reliability coefficient than the groups which experienced only the filmstrip and program (FSP). An additional four computations were significant between .05 and .10 level of probability, one of which went counter to prediction. Using .10 as the largest permissible error, 12 of the 28 calculations were significant and two of those 12 occurred in a direction opposing the predicted outcome.

In general the most effective combination of instructional inputs was the filmstrip and the program (FSP), accounting for five of the 10 significant results in favor of the rationale. These findings seem to be a reasonably good support of the prediction that a systematic instructional in-

TABLE 20

CORRELATIONS BETWEEN PRE AND POST SPACE RELATIONS TEST BY TREATMENT (SEX = MALE)

School	Treatment	N	r	Significance
Ypsilanti	C	18	.901	.01
"	P	32	.793	.01
St. Clair Shores	C	12	.844	.01
"	P	52	.719	.01
"	F,P	38	.763	.01
Walled Lake 8th	C	19	.708	.01
"	P	14	.707	.01
"	F	31	.815	.01
"	F,P	17	.841	.01
"	FS	8	.872	.01
"	FS,P	7	.740	
"	F,FS	6	.868	.05
"	F,FS,P	13	.738	.01
Walled Lake 9th	C	25	.803	.01
"	P	28	.877	.01
"	F	21	.917	.01
"	F,P	30	.821	.01
"	FS,P	15	.822	.01
"	F,FS	24	.901	.01
Saline	C	11	.903	.01
"	F	10	.935	.01
"	F,P	14	.731	.01
"	FS	12	.902	.01
"	F,FS,P	10	.845	.01

Legend:
C = Control
P = Program
F = Film
FS = Filmstrip

All correlations but two were significantly greater than zero at the .01 level of probability.

TABLE 21

CORRELATIONS BETWEEN PRE AND POST SPACE RELATIONS TEST BY TREATMENT
(SEX = FEMALE)

School	Treatment	N	R	Significance
Ypsilanti	C	18	.871	.01
"	P	27	.773	.01
St. Clair Shores	C	19	.787	.01
"	P	39	.719	.01
"	F	26	.907	.01
"	F,P	23	.916	.01
Walled Lake 8th	C	31	.810	.01
"	P	16	.625	.01
"	F	13	.768	.01
"	F,P	8	.921	.01
"	FS	11	.768	.01
"	FS,P	13	.485	
"	F,FS	21	.884	.01
"	F,FS,P	13	.766	.01
Walled Lake 9th	C	18	.861	.01
"	P	18	.691	.01
"	F	20	.823	.01
"	F,P	23	.804	.01
"	FS,P	22	.882	.01
"	F,FS	25	.850	.01
Saline	C	19	.813	.01
"	F	8	.583	
"	F,P	5	.860	
"	FS	4	1.000	.01
"	F,Fs,P	13	.899	.01

Legend:
C = Control
P = Program
F = Film
FS = Filmstrip

All correlations but three were significantly greater than zero at .01 level of probability.

2

TABLE 22

CORRELATIONS AND z-TRANSFORMATIONS UNDER POOLED GROUPS BY TREATMENT AND BY SEX

	SPACE RELATIONS TEST (DAT)			GEOMETRIC INVENTORY (GI)		
	Boys	Girls	Combined Sex	Boys	Girls	Combined Sex
Control	$\bar{z}$ = 1.195 $\bar{r}$ = .83 n = .84	$\bar{z}$ = 1.181 $\bar{r}$ = .83 n = 96	$\bar{z}$ = 1.187 $\bar{r}$ = .83 n = 180	$\bar{z}$ = 1.221 $\bar{r}$ = .84 n = 84	$\bar{z}$ = 1.225 $\bar{r}$ = .84 n = 96	$\bar{z}$ = 1.223 $\bar{r}$ = .84 n = 180
Program	$\bar{z}$ = 1.120 $\bar{r}$ = .80 n = 122	$\bar{z}$ = .904 $\bar{r}$ = .72 n = 95	$\bar{z}$ = 1.026 $\bar{r}$ = .77 n = 217	$\bar{z}$ = 1.111 $\bar{r}$ = .80 n = 122	$\bar{z}$ = 0.720 $\bar{r}$ = .61 n = 95	$\bar{z}$ = 943 $\bar{r}$ = .73 n = 217
Film	$\bar{z}$ = 1.361 $\bar{r}$ = .87 n = 57	$\bar{z}$ = 1.237 $\bar{r}$ = .84 n = 66	$\bar{z}$ = 1.298 $\bar{r}$ = .86 n = 123	$\bar{z}$ = 0.955 $\bar{r}$ = .74 n = 57	$\bar{z}$ = 0.882 $\bar{r}$ = .70 n = 66	$\bar{z}$ = .916 $\bar{r}$ = .72 n = 123
Filmstrip	$\bar{z}$ = 1.432 $\bar{r}$ = .89 n = 19	$\bar{z}$ = 1.235 $\bar{r}$ = .84 n = 15	$\bar{z}$ = 1.355 $\bar{r}$ = .87 n = 34	$\bar{z}$ = 0.699 $\bar{r}$ = .60 n = 22	$\bar{z}$ = .908 $\bar{r}$ = .71 n = 15	$\bar{z}$ = .784 $\bar{r}$ = .65 n = 34
Film & Program	$\bar{z}$ = 1.078 $\bar{r}$ = .79 n = 92	$\bar{z}$ = 1.363 $\bar{r}$ = .87 n = 57	$\bar{z}$ = 1.178 $\bar{r}$ = .82 n = 149	$\bar{z}$ = 0.861 $\bar{r}$ = .69 n = 92	$\bar{z}$ = 1.023 $\bar{r}$ = .84 n = 57	$\bar{z}$ = .921 $\bar{r}$ = .72 n = 149
Filmstrip & Program	$\bar{z}$ = .950 $\bar{r}$ = .74 n = 6	$\bar{z}$ = .530 $\bar{r}$ = .48 n = 12	$\bar{z}$ = 0.650 $\bar{r}$ = .57 n = 18	$\bar{z}$ = 1.658 $\bar{r}$ = .925 n = 6	$\bar{z}$ = 1.018 $\bar{r}$ = .76 n = 12	$\bar{z}$ = 1.242 $\bar{r}$ = .84 n = 18
Film & Filmstrip	$\bar{z}$ = 1.459 $\bar{r}$ = .89 n = 31	$\bar{z}$ = 1.318 $\bar{r}$ = .86 n = 45	$\bar{z}$ = 1.371 $\bar{r}$ = .87 n = 76	$\bar{z}$ = 1.046 $\bar{r}$ = .77 n = 31	$\bar{z}$ = .809 $\bar{r}$ = .66 n = 45	$\bar{z}$ = .899 $\bar{r}$ = .71 n = 76
Film, Filmstrip, & Program	$\bar{z}$ = 1.106 $\bar{r}$ = .80 n = 38	$\bar{z}$ = 1.310 $\bar{r}$ = .86 n = 46	$\bar{z}$ = 1.223 $\bar{r}$ = .84 n = 84	$\bar{z}$ = 1.441 $\bar{r}$ = .89 n = 38	$\bar{z}$ = .847 $\bar{r}$ = .68 n = 46	$\bar{z}$ = 1.088 $\bar{r}$ = .79 n = 84

TABLE 23

SIGNIFICANT DIFFERENCES AND LEVELS BETWEEN CORRELATIONS OF GROUPS POOLED BY TREATMENT, SEX AND SCHOOL. SPACE RELATIONS TEST.

GROUP:	C	P	F	FS	FP	FSP	FFS	FFSP
C	-	.08				.05		
P		-	.05	.06		.10	.05	.09*
F			-			.05		
FS				-		.05		
FP					-	.05		
FSP						-	.05	.05*
FFS							-	
FFSP								-

Legend:

C = Control
P = Program
F = Film
FS = Filmstrip
* = results occurred counter to prediction

put inserted between pre- and posttests would result in a significant reduction of the reliability coefficient when the control group was used as the principal basis of comparison. More generally, however, interpretation of results given in Table 23 can be clarified by stress on instructional loading. From the greatest to the least instructional concentration by a single component was as follows: program, filmstrip, film, and control. The program represented an effort to provide a large amount of detailed information to help Ss deal with problems of development while the control group received no special instruction. The filmstrip was the next most concentrated input while the film was used only as a motivational device. Consequently, a rank order of the treatments, both in combination and singly on the basis of instructional loading, can be represented as follows.

Treatment	Rank Order
P,FS,F	1
P, FS	2
P, F	3
P	4
FS, F	5
FS	6
F	7
C	8

The program, filmstrip, and film combination was intended to represent the most intensive and comprehensive instructional effort. The program plus the filmstrip was intended as the second strongest input, followed by the program and film combination and so on to the control group, which received no special instruction.

According to the above rank orders of treatments, two of 12 correlational differences, shown in Table 23, occurred counter to expectancy. The film-filmstrip-program group was exceeded by the program group.

While the first hypothesis cannot be rejected when calculations were based on individual groups divided according to school, sex, and classroom, there is a reasonable basis for interpreting the results as favorable to the rationale when groups were pooled according to treatment, sex, and school. Pooling, of course, increased N's and consequently reduced the standard error of the distribution of differences.

When data were combined by treatment and by school and when differences between transformed r's (into z s) were computed according to sex division, none of the differences calculated on male data were significant at the .05 level of probability. Four differences, however, were significant at the

.10 level, all in the direction favorable to prediction. Female Ss supplied data containing most of the significant differences. Eleven of the possible 28 differences were significant at .10 level or better, based on female data. Eight of the 11 differences were at the .05 level or better; and three of the 11 were counter to expectancy. In general, females performed more nearly in accordance with our rationale than did the males in respect to the first hypothesis.

THE SECOND HYPOTHESIS

In the null form the second hypothesis predicted that no significant differences would be found between group means (Space Relations Test) after differential treatments had been experienced. The rationale indicated a rejection of the null hypothesis on the ground that instructional input, especially via program and filmstrip, would result in learning that would transfer to the post Space Relations Test. Because the program and filmstrip represented efforts to inform students how to deal with problems of development and because the Space Relations Test was intended to measure the development aspect of space perception, mean scores of those groups which used the program and the filmstrip were expected to be significantly greater than those groups which received no instruction or experienced only the film which was limited to motivating students rather than providing cognitive input regarding the solutions to problems.

Table 24 indicates the results according to the covariance analysis for boys only. In each of the four schools all groups receiving the same treatment were pooled and F-ratios were obtained for each school grouping. In the Walled Lake School, however, ninth graders were grouped separately; thus, five covariance analyses were run for each sex division.

One advantage of the covariance technique used in this study can be credited to the efforts of Professor Clemens Johnson of the Computing Center and the School of Education, The University of Michigan. Dr. Johnson developed a technique for analyzing data by covariance while using unequal N's in the various groups. Hence, no data were discarded to equalize N's. Dr. Johnson also developed a program for the computer so that the covariance analysis could be handled by the available computing equipment.

Table 24 includes the adjusted means for each of the sub-groups within the school divisions and the computed F-ratios. Only two of the five F-ratios were significant for boys at the .01 level of probability. The Ypsilanti school and the ninth grade at Walled Lake supplied the significant results.

The t-ratio between control and program groups at Ypsilanti was significant at the .01 level of probability. No other groups except program and control sections were in the Ypsilanti sample, the pilot group.

TABLE 24

ADJUSTED MEAN SCORES BY COVARIANCE ANALYSIS FOR MALES BY TREATMENT WITHIN SCHOOLS (SPACE RELATIONS TEST)

School	Treatment	N	Adj. M	F-ratio	Mean Gain
Ypsilanti	P	31	50.14	7.89*	13.32
	C	18	39.37		1.78
St. Clair Shores	P	52	56.30	.35	5.77
	F,P	35	55.16		6.57
	C	12	51.99		1.67
Walled Lake 8th Grade	FS,P	6	57.28	1.33	17.83
	FS	9	56.81		16.66
	F,FS,P	13	53.64		14.38
	P	13	50.35		10.38
	F,P	15	49.75		10.33
	F,FS	6	48.25		5.67
	C	17	48.15		8.41
	F	27	43.70		3.41
Walled Lake 9th Grade	F,P	28	60.59	6.86*	17.68
	P	26	60.10		17.39
	F,FS,P	15	53.76		9.60
	C	26	50.01		6.74
	F	20	47.72		4.20
	F,FS	25	45.27		2.20
Saline	F	10	66.90	2.26	18.60
	C	11	65.13		12.64
	F,FS,P	10	63.53		11.50
	F,P	14	61.80		12.00
	FS	10	52.61		3.70

Legend: C = control
F = film
FS = filmstrip
P = program
all others are combination treatments of those listed above

*significant at .01 P

Among the six groups in the ninth grade, Walled Lake School, the program received a significantly higher mean on the criterion test than the control group. The t-value was significant at the .01 level. Also, the combination of film and program resulted in a mean significantly higher than the control group.

In general, the data supplied by boys supported our expectation only partially. The boys at the Saline school provided data almost in direct opposition to that which was expected according to the rationale. The control and film groups received the highest means while the combination of film, filmstrip, and program was third in rank order. In all other schools, group mean rank orders were fairly consistent with prediction although only two of the five F-ratios were significant. The selected treatments for boys had a significant effect in two of the five large groups.

Table 25 presents comparable data for girls. Again, only two of the five F-ratios were significant, both at the .01 level of probability. In both cases S's who used the program either singly or in combination with other instructional components received significantly higher means on the criterion test than S's in the Control Group.

In the Walled Lake school, ninth grade, three of the possible five t-ratios were significant at the .05 level or better when the Control Group was the basis for comparison. The combination of film and filmstrip as well as the film alone failed to produce results significantly greater than the Control Group.

The F-ratio for Ypsilanti girls was 3.96, which approached the critical value of 4.08 for significance at the .05 level of probability.

St. Clair Shores, Walled Lake (eighth grade), and Saline schools produced results which failed to support the rationale. While the data supplied by girls agreed slightly more with the expected outcomes than data supplied by boys, the support was only modest and failed to represent a clear-cut superiority of instructed groups over those which received no instruction.

High and Low Intelligence.

On the basis of mental age, S's were grouped in two divisions: "high" and "low." A covariance analysis was performed on each division by school, using the means of the control groups as the baseline for comparison. In each instance the program group was compared with the control except in the Saline school where the program alone was not used. There the experimental group had experienced the program, film, and filmstrip combination. In addition a sixth group was identified as "Transients" or S's who had a record of a relatively high number of school transfers. The "Transient" group was identified because the level of achievement of such S's can be reasonably

TABLE 25

ADJUSTED MEAN SCORES BY COVARIANCE ANALYSIS FOR FEMALES BY TREATMENT WITHIN SCHOOLS (SPACE RELATIONS TEST)

School	Treatment	N	Adj. M	F-ratio	Mean Gain
Ypsilanti	P	27	46.72	3.96	15.63
	C	18	37.70		6.50
St. Clair Shores	P	37	58.62	8.92*	15.51
	F,P	23	48.91		6.30
	C	16	46.52		2.78
	F	28	43.71		1.88
Walled Lake 8th Grade	P	16	63.98	1.94	21.44
	F,FS,P	11	56.94		15.09
	F,P	7	56.82		14.72
	F,FS	21	55.43		10.47
	C	31	52.03		10.06
	FS	11	49.49		9.28
	F	14	48.93		7.43
	FS,P	12	48.00		7.84
Walled Lake 9th Grade	P	15	66.02	5.45*	23.47
	F,P	22	60.12		17.00
	F,FS,P	22	56.72		10.86
	F	19	54.24		11.05
	F,FS	24	50.34		7.75
	C	14	46.62		4.36
Saline	F,FS,P	13	64.75	1.75	20.08
	C	17	64.62		18.29
	F,P	5	63.70		16.60
	F	5	59.10		15.00
	FS	4	49.90		6.50

Legend: C = control
F = film
FS = filmstrip
P = program
all others are combination treatments of those listed above

*significant at .01 P

related to the number of new adjustments they are called upon to make. S's in the low mental age grouping could be expected to adjust more slowly to new academic situations than those of higher mental age when CA is held constant. Programmed materials which are liberally cued to encourage desired responses are likely to be especially advantageous for S's of low I.Q. If so, it is logical to predict that the low MA group which had received programmed instruction would profit relatively more than S's in the high MA group.

Table 26 shows results of the High MA Group, with adjusted means on post Space Relations Test, by school and treatment. F-ratios were significant in two cases of the six, namely, Ypsilanti and Walled Lake ninth grade. Ypsilanti S's in the program group gained about three times more than the control group in that school, while in the Walled Lake ninth grade the program group gained about twice as much as the control. Although all adjusted means in the other groups favor those who experienced the program over the control groups, differences are not significant.

TABLE 26

COVARIANCE ANALYSIS OF PRE- AND POST- SPACE RELATIONS TEST SCORES
FOR SUBJECTS GROUPED AS HIGH MA
(Mean I.Q. 110)

School	Treatment	Adjusted Mean	F	df	Mean Gain	N
Ypsilanti	P	69.66	5.16**	28	23.20	15
	C*	59.88			7.73	15
St. Clair	P	70.76	.99	62	9.46	46
Shores	C	66.85			4.33	18
Walled Lake	P	64.31	1.28	29	17.09	12
8th Grade	C*	58.32			10.37	19
Walled Lake	P	63.87	7.11**	33	17.30	17
9th Grade	C	53.68			7.94	18
Saline	P,FS,F	76.92	.17	14	18.09	11
	C	75.58			12.20	5
Transients	P	61.79	1.02	15	9.88	8
	C	56.18			3.89	9

*Pre-control higher than pre-program
**Significant at .05 level or better, Transients were members of all schools

In Table 27 the same kinds of data for the Low MA Group are presented. Three of the six F-ratios are significant at the .05 level or better. In all cases except the Saline groups, the mean gains for the controls were from almost minus-three points to nearly nine while mean gains for the experimental groups were from a low of about 10 to a high of over 27. The most significant difference was found with Transients—the program group gaining over 10 times that of the control group.

TABLE 27

COVARIANCE ANALYSIS OF PRE- AND POST- SPACE RELATIONS TEST SCORES FOR SUBJECTS GROUPED AS LOW MA

(Mean I.Q.: 94)

School	Treatment	Adjustment Mean	F	df	Mean Gain	N
Ypsilanti	P	35.96	4.75**	53	10.82	39
	C	28.98			2.87	16
St. Clari	P	50.74	5.80**	51	10.21	43
Shores	C	37.32			-2.80	10
Walled Lake	P	48.42	2.17	36	14.62	13
8th Grade	C	41.73			7.00	25
Walled Lake	P	52.12	2.74	19	19.70	10
9th Grade	C	42.08*			8.73	11
Saline	P,FS,F	65.44	.80	20	14.13	8
	C	68.96*			18.93	14
Transients	P	50.75	10.16**	22	27.28	14
	C	24.36*			2.30	10

*Pre-control higher than pre-program

**Significant at .05 level or better

Transients were members of all schools

Although the analysis of gain scores is difficult for a variety of technical reasons, the aim of instruction is to promote increments in measures of achievement. Without attempting any statistical interpretation of gains for control and experimental groups of high and low MA divisions, it is worthwhile to note them in passing. The following estimates of median gains suggest certain speculations that will be indicated in the following chapter:

	High MA	Low MA
P	17	14.50
C	8	5.50

The above figures indicate that the Low MA group under programmed instruction gained considerably more than the High MA group which received no instruction. While the actual means for the High MA Control Groups were higher than the Low MA Experimental Groups (P), results indicate that educational inputs tend to increase criterion scores for low MA S's more than high MA S's when calculated on percentage gain; yet absolute gains favor the high MA group. These results are consistent with many other comparisons between gains of high and low MA groups.

Rank Orders of Treatments.

On the basis of criterion score gains, adjusted according to the covariance analysis, Figure 4 shows results of pooled Ss' by treatment. Differences between rank orders indicated in Table 28 and those predicted yielded a Rho of .81, suggesting that instructional inputs had an effect similar to that expected. Table 28 shows expected rank orders versus obtain rank orders.

TABLE 28

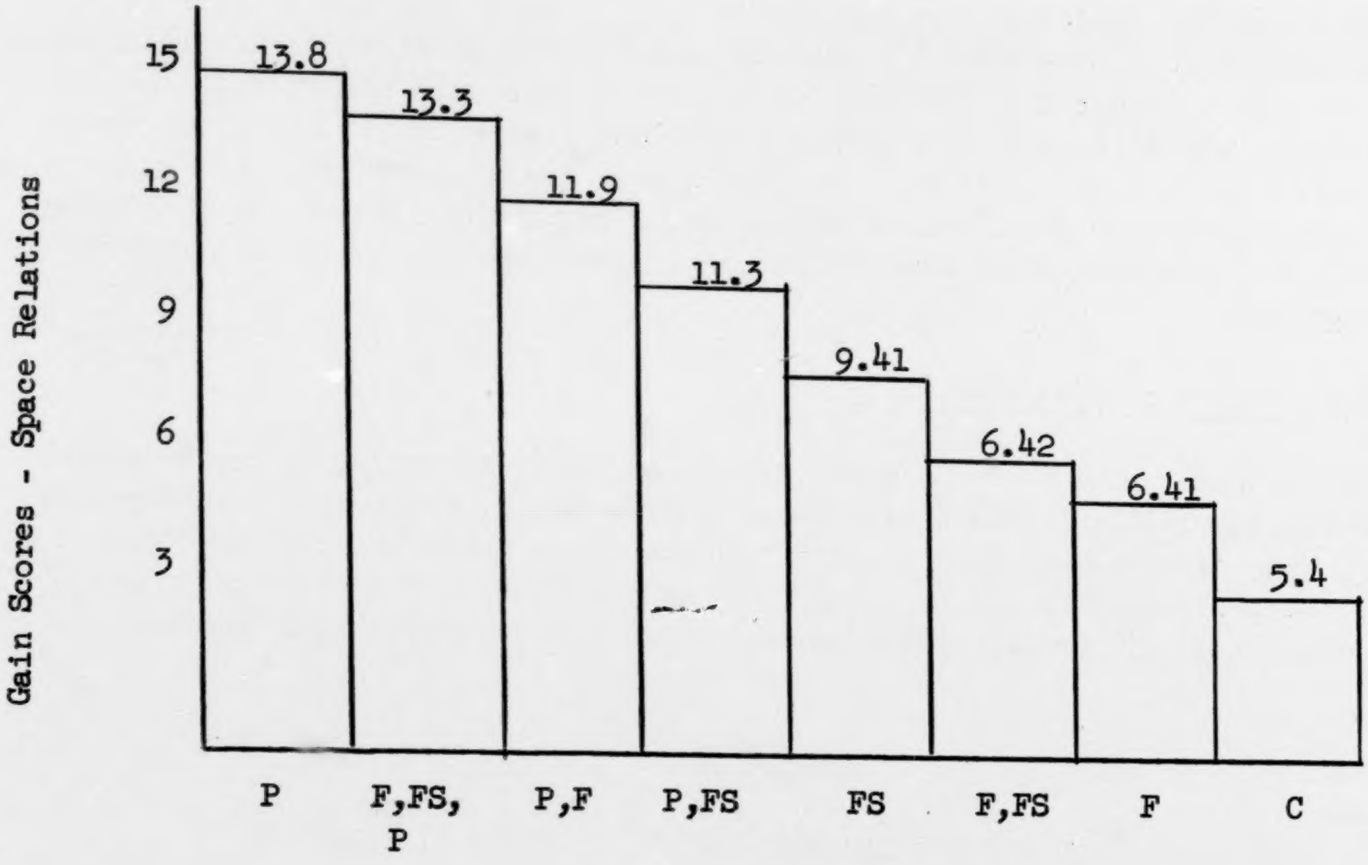
EXPECTED VERSUS OBTAINED RANK ORDERS OF GAINS BY TREATMENT
(Gains on Basis of Adjusted Means)

Treatment	Expected Rank Order	Obtained Rank Order
P,FS,F	1	2
P,FS	2	4
P,F	3	3
P	4	1
FS,F	5	6
FS	6	5
F	7	7
P	8	8

Rho = .81

FIGURE 4

ADJUSTED MEAN GAINS BY TREATMENT AND BY POOLED SUBJECTS



Treatment

Legend:

P = program
F,FS,P = film, filmstrip, program combination
P,F = program and film combination
P,FS = program and filmstrip combination
FS = filmstrip
F,FS = film and filmstrip combination
F = film
C = control

Tables 29 and 30 summarize the results concerning means adjusted by the covariance analysis on the Geometric Inventory Test according to treatments within each school. As mentioned before, the Geometric Inventory was a special achievement test developed to assess learning of the programmed content only. The inventory was not geared to any other input. Table 29 shows that all groups which received the program gained between 14 and 25 points while the control groups gained from about three to slightly over five points. The large F-ratios reflect the significance of differences between means of those groups which took the program and means obtained by groups that did not experience the program.

The consistency of gain scores by groups that received the program is rather high from school to school. The consistency of low gains made by non-program groups was even greater. While groups performed as expected on the Geometric Inventory the amount of transfer to the Space Relations Test (DAT) was less than expected. These results suggest that the program did not provide special practice *per se* for taking the Space Relations Test.

TABLE 29

ADJUSTED MEAN SCORES BY COVARIANCE ANALYSIS FOR MALES BY TREATMENT WITHIN SCHOOLS (GEOMETRIC INVENTORY TEST)

School	Treatment	N	Adj. M	F-ratio	Mean Gain
Ypsilanti	P	31	46.62	69.65	22.55
	C	18	22.27		3.17
St. Clair Shores	P	52	50.85	21.44	20.03
	F,P	35	47.18		17.09
	C	12	33.80		2.67
Walled Lake 8th Grade	P	13	46.23	27.98	22.08
	FS,P	6	46.15		22.30
	F,P	15	44.12		19.40
	F,FS,P	13	39.70		15.85
	FS	9	29.32		5.56
	C	17	29.12		4.30
	F	27	27.63		3.37
	F,FS	6	22.48		1.83
Walled Lake 9th Grade	F,FS,P	15	49.49	30.16	22.00
	F,P	28	43.62		16.18
	P	26	41.70		14.27
	C	26	31.40		3.96
	F,FS	25	30.69		3.24
	F	20	28.82		1.35
Saline	F,FS,P	10	59.56	33.27	25.40
	F,P	14	50.17		16.07
	C	11	45.77		5.28
	F	10	30.88		2.40

Legend: C = control
F = film
FS = filmstrip
P = program
all others are combination treatments of those listed above

Note: All F-ratios significant beyond .01 P

TABLE 30

ADJUSTED MEAN SCORES BY COVARIANCE ANALYSIS FOR FEMALES BY TREATMENT WITHIN SCHOOLS (GEOMETRIC INVENTORY TEST)

School	Treatment	N	Adj. M	F-ratio	Mean Gain
Ypsilanti	P	27	42.82	36.29	24.04
	C	18	22.93		2.94
St. Clair Shores	P	37	53.59	108.27	21.89
	F,P	23	52.97		21.78
	C	16	35.17		3.00
	F	28	34.54		5.00
Walled Lake 8th Grade	P	16	52.17	40.78	23.63
	F,P	7	48.97		20.43
	FS,P	12	47.55		19.00
	F,FS,P	11	45.64		17.09
	FS	11	33.73		5.18
	C	31	30.93		2.38
	F	14	30.69		2.14
	F,FS	21	30.25		1.72
Walled Lake 9th Grade	F,FS,P	22	51.53	45.48	21.64
	P	15	50.96		23.20
	F,P	22	49.25		21.50
	F,FS	24	32.36		4.54
	F	19	32.18		4.00
	C	14	31.38		3.79
Saline	F,FS,P	13	58.60	23.83	23.84
	F,P	5	51.69		12.80
	C	17	45.55		3.64
	FS	4	42.47		8.50
	F	5	34.31		3.40

Legend: C = control
F = film
FS = filmstrip
P = program
All others are combination treatments of those listed above

Note: All F-ratios significant beyond .01 P

CHAPTER FIVE: DISCUSSION

This study was an effort to assess the stability of a standardized measure of space relations under conditions of varying instructional inputs which intervened between pre- and posttests. The second objective was to determine the relative effects of the instructional components as measured by the criterion test of space relations. The rationale pertaining to these efforts is recapitulated as follows:

a. It is not reasonable to expect that each child has had an equal opportunity in his background of experience to approach the taking of a measure of intelligence or aptitude so that the competition is equally fair for all. It is reasonable, however, to expect that children differ widely in their backgrounds and consequently represent a wide distribution of learning opportunities appropriate to the test. This is particularly likely for measures that represent dimensions not found in the common core of the curriculum. No systematic instruction is given to students in dealing with the various kinds of problems of space relations. With the possible exception of art training in higher education, subject matters which require the ability to deal with problems of space make no special provisions to improve that ability. In solid geometry, for example, stress is placed on the proof of theorems through deduction, although the ability to visualize space relations is a significant factor for success in that subject. Instead of training in space relations, students are assumed to have the "space component of ability" already developed as a requisite. If, however, special instruction on how to handle problems of space can produce significant gains, it would seem that some part of the curriculum should be used for such teaching. The argument becomes even more convincing when it is seen that the ability in dealing with space relations is apparently a significant factor in a number of other areas, including mechanical drawing, shop mechanics, dress making, various courses in engineering, architecture, dentistry, and even some aspects of physics, chemistry, and biology.

Perhaps one reason why no systematic effort is made to teach students how to deal with space problems is that the ability is assumed to resist education because it is a product of genetic factors. That reason, however, may be no better than the argument for eliminating reading, writing, spelling, arithmetic and other subjects because it seems that I.Q. is a significant determiner of academic success. Because no psychological trait has been definitively found to be solely the product of heredity despite the assumptions which lean in that direction regarding factors of intelligence, it would seem that the composition of the curriculum should be based on the importance of the knowledge or skill rather than upon assumptions which are not supported by known facts. Data which bear upon academic achievement suggest that success in all subjects is at least partially a function of interaction of both nature and nurture. Also, the fact that no sufficient assessment has been made to account for success in dealing with space relations leaves us with the notion that both genetics and experience are significant determinants.

Predictions of academic success based on I.Q. and aptitude measures involve a broad band of error. In some samples the error of prediction, based upon accounted-for variance, may be as low as 30 per cent. But in other samples the error factor is above 80 per cent. It is usually not an easy matter to partition the total reliability error into clear-cut components. Systematic influences can sometimes be confused with chance error. Learning is one such influence. The probable confusion between the effects of learning and random error on reliability measures of aptitudes is fundamental in the present rationale. Assuming that the upper limit of achievement is controlled by the existence of native capacity, represented by a broad band of values in the population, and assuming that Ss differ considerably in terms of developed capacity, any sample of aptitude measures will represent different degrees of developed capacity. In the case of the space relations aptitude, which is not developed systematically by curricular efforts, cultivation of the aptitude is left to fortuitous circumstances. It is therefore reasonable to suspect that some Ss have much more of this native capacity developed than others. Consequently, Ss with low development but with high capacity will gain considerably more than Ss with medium or high development but having only medium or low capacity. One major effect of systematic instruction in space relations would be a significant shifting of individual rank orders between pre- and posttests. This rationale was used to support the expectancies in the testing of the first hypothesis, which indicated that instruction inserted between pre- and posttests would lower the correlation values significantly when compared with correlations of scores furnished by control Ss who received no instruction.

b. The second hypothesis dealt with the effects of the various instructional treatments. The program was the single input having the greatest concentration of informative content. The filmstrip had a modest degree of specific information while the film had the least. The film was intended only to motivate Ss to make a maximum effort on the criterion tests by illustrating the importance of the ability to deal with space relations. A combination of all three inputs was expected to yield the greatest achievement, followed by the filmstrip-program combination and so on to zero instruction given to the control group. A rank order correlation between expected and obtained results, using adjusted mean posttest scores as measures, was intended as a gross index for judging the prediction. A covariance analysis was used to assess the influence of treatments by schools, grade, and sex.

As indicated in the preceding chapter, results used to test the first hypothesis (that instruction between pre- and posttests would lower reliability correlations) were not a clear-cut support of the prediction. One significant difficulty that probably contributed to the lack of unamimous support by data sorted into groups was imperfections in the criterion test. Under the time and monetary limits of this study it was necessary to select a criterion test already in existence. After considerable examination of the available standardized tests of space relations, it was decided that the one in the Differential Aptitude Battery was the most satisfactory. Yet, results showed that reliability coefficients based on control groups (which received no instruction) varied from a low of .708 to a high of .903. It therefore seems that the stability of the selected test was not optimal for carrying out the plan of this research. The reliability coefficient of .708 leaves about 50 per cent of the variance unaccounted for - suggesting that chance error is greater than one would suspect by reading the DAT manual. It can be argued, however, that if the rationale above is reasonably correct, the effect of instruction should have lowered the reliability coefficients more than indicated by the results.

A second possible weakness lies in the instructional inputs. It is extremely doubtful that the treatments represented optimal teaching and learning for the purpose intended. Although much time and effort were given to the development of the programmed content, its relevance to the intended purpose can be debated. Results of the Geometric Inventory based upon the program indicated overwhelming gains in favor of Ss who received the program. Gains of other groups were insignificant.

A third probable weakness in the study is that conditions for transfer were less than adequate. If transfer can be considered a function of common elements between two sets of experiences, face validity of the Geometric Inventory and the Space Relations Test suggests the kind of overlapping that approaches training for the criterion test. How similar should the learning experience be to the test for transfer? The answer is a difficult one. Some examiners of the program may charge that some frames are too much like the items on the Space Relations Test. But results cannot support that charge unequivocally.[1] If instruction has amounted to special training for the test, differences between means (of program and non-program groups) would have been more striking than those obtained.

A fourth factor which could have contributed to diminution of strength of the independent variables was the teacher variable. An effort was made to estimate the attitude of teachers toward the use of the program. The many difficulties involved in getting valid and reliable measures of attitudes influenced the decision of minimizing dependence upon elaborate indices of attitude. Although teachers were paid a stipend to include the program in their courses on mathematics, their attitude toward the task was, in general, less than enthusiastic. The problem of persuading teachers and administrators to alter their schedule to include experimental operations is always a delicate one. Several meetings were held with teachers after initial contacts had been made with the principal and teachers in each school. The main purpose of those meetings was to establish rapport with the teachers and to promote a cooperative and positive attitude toward the study. While it is believed that the results of those efforts were better than usually found in studies requiring the cooperation of classroom teachers, the fact that certain rules must be followed is likely to be felt by the teacher as limiting his autonomy. Perhaps the only ideal solution to the problem lies in the development of an experimental school where teachers are selected that are able and willing to commit themselves to research in education.

---

[1] A supplementary study of this question is now in progress.

Despite the combination of many probable effects that dissipated the strength of the experimental treatments, results pertaining to the first hypothesis indicated more-than-chance affirmation. It is extremely improbable that measures selected at random could have produced 10 significant differences between transformed correlation coefficients from a possible total of 28. On the basis of chance alone less than three significant differences could be expected, using .10 as the largest permissible error. The program and filmstrip combination produced five significant differences when compared with inputs of lesser intensity. In other words, the combined groups which experienced both the program and filmstrip resulted in data that reduced the stability of predictions based on the Space Relations pretest. Two of the 12 significant differences were counter to expectancy. Yet, even those comparisons favored groups which experienced the program. The combined control groups produced no results superior to any of the programmed groups in terms of the first hypothesis.

The general significance of results used to assess the first hypothesis is far from definitive because of sampling limitations. Yet, the important implication is that measures of the space relations aptitude should be used with extreme caution. The second implication is that the touted reliability of the Space Relation Test (DAT) may be an artifact produced by gaps in the curriculum. The stability of other aptitude tests which are more adequately represented by curricular experiences may be an artifact of teaching-learning deficiencies. While these points are only speculations they suggest that emphasis upon measurement can be misleading. One common practice in the use of psychological measurements in education is the sorting of students into various classifications. Once a student has been "typed" by his profile on a battery of tests there is probably a tendency on the part of teachers and administrators to form certain expectancies in agreement with the "typing." These expectancies are likely to predispose the establishment of a reinforcement system which operates to strengthen behavior patterns consistent with the expectancies. While the net effect of such likely practice is not all negative, the probable result is the shaping of a stereotype. And as the student learns to adjust to these expectancies development of his creative abilities appears to diminish.

From this study and others it would seem that the most adequate basis for making predictions of academic success can come only after the student has experienced conditions of learning that approach an optimal level. Measures of performance under those conditions would probably be the best indication of future achievement. This viewpoint suggests that an increased emphasis be placed on problems of teaching-learning management and a decreased emphasis upon current measures of aptitude and intelligence. If the influence of the school is limited to a special environmental impact, the matter of prime importance is the shaping of an adequate set of conditions for learning.

Traditional education seems to imply a two factor theory. According to such theory academic achievement is determined by genetic endowment and by opportunity. Therefore, a child who has inherited good genes for learning needs only the availability of materials, apparatus, and teachers who are willing to impart their knowledge and skill. Students who fail under the system are deemed to suffer some deficiency - either mental or emotional, or both. Emphasis is upon adjustment to the status quo. While the importance of adjustment to prevailing conditions is important, the aim of education is to help students develop as much as possible productive expressions of their capacities. Consequently, it is logical to assume that school environments should be made flexible to provide proper nurture for those who find it difficult to learn under a set of narrow restrictions. The logical outcome would seem to be an emphasis upon both adjustment to and adjustment of the school environment. Rigid adherence to narrow interpretations of psychological measurements could act to support unduly a stress upon the first aspect of adjustment.

A theory of intelligence which broadens the base of environmental factors of influence seems to be implied by the results of this study. If it is assumed that aptitudes are factors of intelligence which require special conditions for their development as determined by the personality pattern of the individual, proper allowance would be made for the kind of flexibility needed in interpreting such measures. More specifically, knowledge of adjustment to a given set of conditions would be required before an adequate meaning could be placed upon a battery of test scores. Such flexibility would serve to give the examinee the benefit of any reasonable doubt, which implies an ethical responsibility for suspending judgment until sufficient evidence is provided. An important possible outcome of this viewpoint would the utilization of psychological measures for helping to define educational problems that are correctible rather than focusing upon presumed limitations of the examinee.

Results on the assessment of the second hypothesis, which predicted that instruction would have no significant effect on criterion scores, suggest equivocal interpretations. On the conservative side it can be indicated that only four of the possible 12 F-ratios were significant the .05 level or better. (See Tables 24 and 25). Programmed instruction therefore failed to indicate its superiority over no instruction in more than 50 per cent of the cases. On the other hand, it is reasonable to regard the several sources of error as working to attenuate the strength of the experimental variables; but that despite the attenuation results favored the experimental groups in more-than-chance frequency.

A reasonably conservative estimate of the significance of this study would support the argument that the interpretation of Space Relations Test scores should be done with extreme caution. It also supports the point that further investigation is needed to estimate the effect of instruction on various aptitude measures.

Perhaps the most worthwhile finding concerned students classed as "transients." When records indicated that a student had changed schools more than three times up to the eighth and ninth grades he was put into the "transient" category. It seems particularly important that transients of low MA who experienced the program gained nearly 12 times more than Ss with low MA who received no instruction. The practical suggestion is that teachers who receive low MA transients may find it profitable to place before such Ss content that is well structured and has ample cueing and feedback provisions such as found in many programmed materials.

The study also suggests that the variable of intensity of instructional input be investigated. Findings in this research effort showed that the most intense combination (film-filmstrip-program) yielded slightly less gains than the program alone. When does a teacher actually over-teach? The study of time distribution of the student during school hours may yield valuable information that may contribute to establishment of optimal proportions of the various activities to the total time. It is likely that individual differences again would be great enough to suggest that all Ss do not profit optimally from a standardized time distribution of activities. Investigations in educational technology where self-instruction is emphasized and individual students move through the curriculum at different paces promise to bring some enlightment about the matter.

It should be recognized that perhaps most experiments in education, including the one reported herein, cannot qualify as true experiments according to the criteria offered by Fisher. When only classrooms are provided on the basis of administrative convenience the criterion of random selection of Ss can hardly be satisfied. Because of this common limitation generalization should await the collection of data of repeated experiments. Unfortunately, adequate duplications are far too rare.

The most urgently needed information that can be supplied by educational research is knowledge about the process of human learning. The present study along with most other experiments in education is designed to examine outcomes rather than changes during the course of learning. Until, however, adequate descriptions of process are available interpretations of studies of the present kind can be only very coarse-grained. This research report, however, can be used to suggest the need for several investigations of learning procedures. The following questions point up some of the obvious research implications.

a. If the student were given the opportunity to structure his experiences in the learning of space relations, what sequential pattern would he adopt? Self-structure could be considered in the way Mager reported it.

b. If a branching program were used to represent the most frequent modes of self-construction, as found in pursuing (a), above, how would the time curve of progress compare with that found by using the linear program of the present study?

c. During the showing of a film designed to motivate Ss, what kinds of pictured events result in significant responses of arousal? What properties of the stimulus complex are correlated with significant arousal? How do these properties differ from those suggested by certain psychological concepts? (Measures of arousal could include the GSR).

d. Using two sets of instructional materials that differ significantly in effectiveness of instruction, how much do the two sets of content differ in evoking arousal? (Programs could be used here of about equal length, but differing to the degree that they meet the criteria of adequate programming.) Does the effective program produce more arousal than the less effective one?

e. Do reinforcing incidents in a learning sequence produce arousal?

f. What special effects do relevant cognitive stimuli that produce positive arousal (plesant feelings) have upon learning? What effects do relevant cognitive stimuli that produce negative feelings have upon learning? (Both GSR and semantic differential measures could be used.)

The above questions point up only a small sample of queries that merit study of the learning process. Adequate answers to such questions could provide an improved knowledge-base for experiments concerned only with outcomes. These considerations lead the write to believe that studies of the present kind are perhaps most valuable for suggesting what investigations of the learning process ought to be performed. It is believed that the major emphasis should be placed on research designed to describe the dynamics of human learning rather than upon studies of outcomes in which details about process can be estimated only by speculation.

CHAPTER SIX: SUMMARY AND CONCLUSIONS

The ability to perceive space relations and to solve spatial problems is tested by a variety of standardized measures. Some tests measure the ability to manipulate mentally flat patterns to develop the proper solid objects. Other tests assess different aspects, namely: ability to integrate or bring together separate pieces to form a whole unit; ability to visualize the perspective of an object after it has been moved along a certain path; and the ability to visualize how a solid will appear after it has been rotated or tilted a certain amount. The present study deals only with the first aspect, the ability to develop flat patterns into solids.

An experiment was designed to evaluate two hypotheses:

a. instructional inputs which intervene between pre- and posttests will have no effect upon test reliability

b. the number and kind of instructional inputs which intervene between pre- and posttests will have no significant effect on altering average or mean scores.

A rationale was used to indicate why both hypotheses would be rejected by the evidence. The rationale is described in Chapter 1.

Three separate instructional inputs were devised. They were: a linear program of over 500 frames, a filmstrip to depict various stages of development of flat patterns into solids, and a short film in color designed to motivate the student to see the value of the ability to deal successfully with problems of space and to motivate him to do his best on the test provided.

The standardized test used in both pre- and posttests was the Space Relations Test of the Differential Aptitude battery. Form A was used for the pretest and Form B was used for the posttest. A second measure was especially constructed to test the effects of the program. This second test was called the "Geometric Inventory."

A total of 881 students supplied usable results for analysis. The students were taken from four different schools in or near Metropolitan Detroit. All Ss were members of the eighth grade except 116 Ss who were ninth graders.

The instruction of Ss involved the following treatments:

a. program
b. film
c. filmstrip
d. program and film
e. program and filmstrip
f. program, film, and filmstrip
g. film and filmstrip
h. control (no instruction given)

Results are summarized below:

a. Concerning the prediction that instruction which intervened between pre- and posttests would have no effect on the reliability coefficient of the test, it was found that 10 differences from a maximum of 28 were significant at the .10 P or better. Seven of the differences were significant at .05 P or better.

b. Results used to evaluate the hypothesis that the kind and amount of instructional inputs would have no effect indicated that four of a maximum of 10 F-ratios were significant at .01 P. No significant reversals to expectation were found.

Using the seven experimental conditions plus the control groups a rank order correlation was computed to compare the gradient of predicted outcomes with actual outcomes. The Rho coefficient was .81, indicating that a fairly high agreement obtained between predicted and actual outcomes.

A number of less than ideal circumstances were met that probably served to weaken the strength of the experimental variables. These weaknesses are treated in Chapter Four.

Perhaps the most significant finding was the fact that Ss with low MA who had changed schools frequently and who received programmed instruction benefitted nearly 12 times more (as measured by gain scores) than their control counterparts.

CONCLUSIONS

1. The results lack a clear-cut consistency in terms of whether or not instruction served to lower reliability coefficients of the Space Relations Test. In 10 cases of a total of 28 comparisons of r's transformed into z's, the data indicated a significant reduction of the reliability coefficient when control groups were used as the baseline.

2. Four of a total of 10 F-ratios were significant at .01 P, indicating that in four experiments out of 10, groups which experienced the programmed material made significantly higher scores on the posttest Space Relations Test than did their control groups. (As shown in Tables 25 and 26, the research can be considered as 10 experiments when Ss were divided according to sex, school, and grade.)

3. All ten F-ratios were significant beyond the .01 P level concerning mean differences between programmed and control groups as measured by the Geometric Inventory.

4. The fact that low MA Ss who were transient Ss and who experienced the program gained nearly 12 times more than low MA Ss on the Space Relations Test (DAT) suggests that such Ss may profit more than most others when given materials in programmed form.

In Chapter 5 several pertinant points were treated on the basis of speculation. Perhaps the most worthwhile speculation dealt with the differences between experiments of outcomes and of processes. It was suggested that outcome experiments may provide their most valuable use in helping to indicate what experiments on learning processes are perhaps most valuable. A sample of such indications was given.

APPENDIX A

ORGANIZATIONAL STRUCTURE FOR PROGRAM DEVELOPMENT

The first step was a description of the behaviors involved in a spatial relations test. The content material to be used in developing these behaviors was then selected from mathematics and was initially assembled in outline form (Appendix A1). This outline was a tentative effort to indicate boundaries for the concepts to be taught in the ultimate program.

In order to show the existence of relationships between the various concepts to be taught, a series of charts were developed (Appendix A2). These schemata helped portray the relationships between concepts and were useful in developing greater coherence within the program.

APPENDIX A1

APPROACH TO PROGRAMMING VISUAL SPACE PERCEPTION

I. Analysis of behaviors to be established.

   A. Criterion behavior: selection of correct response(s) in the form of a solid figure from a series of alternatives with no external stimulus support other than the stimulus image in the form of a flat pattern.

   B. Components of criterion behavior. Greater use of estimation in:

      1. Differentiation or discrimination--perception of togetherness and apartness, shapes, sizes, markings and shadings, etc.

      2. Identification--recognition of familiar object characteristics and attaching appropriate labels.

      3. Organization or relationship--various parts within a given pattern followed through movement path to completely folded solid.

      4. Orientation--in terms of personal orientation to object image.

II. Outline of techniques for developing criterion behavior.

   A. Establish a "set" for differentiations and discrimination of various cues on the basis of estimation.

      1. Develop discriminations of simple cues in isolation--point, set, line, line segment, ray, angle, etc.

      2. Integrate and discriminate cues in 2-d figures.
         a. rectangles
         b. squares
         c. triangles
         d. circles

      3. Differentiate within and between classes.
         a. triangles
         b. quadrilaterals
         c. circles

      4. Integrate pattern of several 2-d component parts into larger 2-d figure.

   B. Provide an introduction to simple solids.

      1. Show representation of solids on plane surfaces.

      2. Show appearance of common types of solids from various angles.
         a. Drawings of solids.
         b. Pattern folding and object manipulation.

C. Show characteristics of various solids in flat pattern (unfolded) form.

    1. Number of sides.

    2. Shape of sides and relative position.

    3. Relative dimension of parts.

D. Present patterns of common solids to be identified (folded and unfolded) and differentiate between classes.

    1. Cubes.

    2. Rectangular solids.

    3. Pyramids.

    4. Cones.

    5. Cylinders.

E. Present patterns of common solids and provide for differentiations within classes.

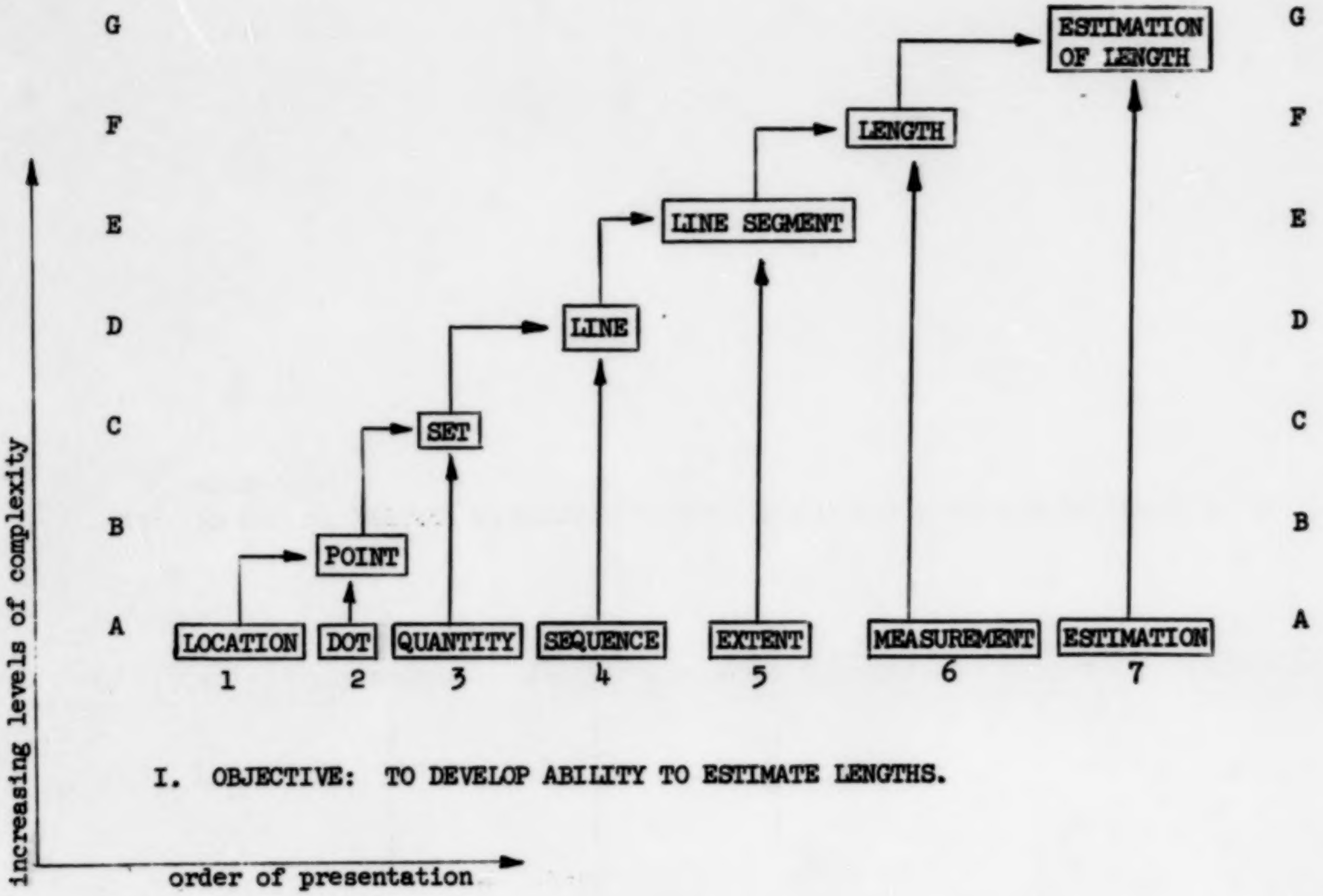
    1. Size differentiations.

    2. Marking or shading differentiations.

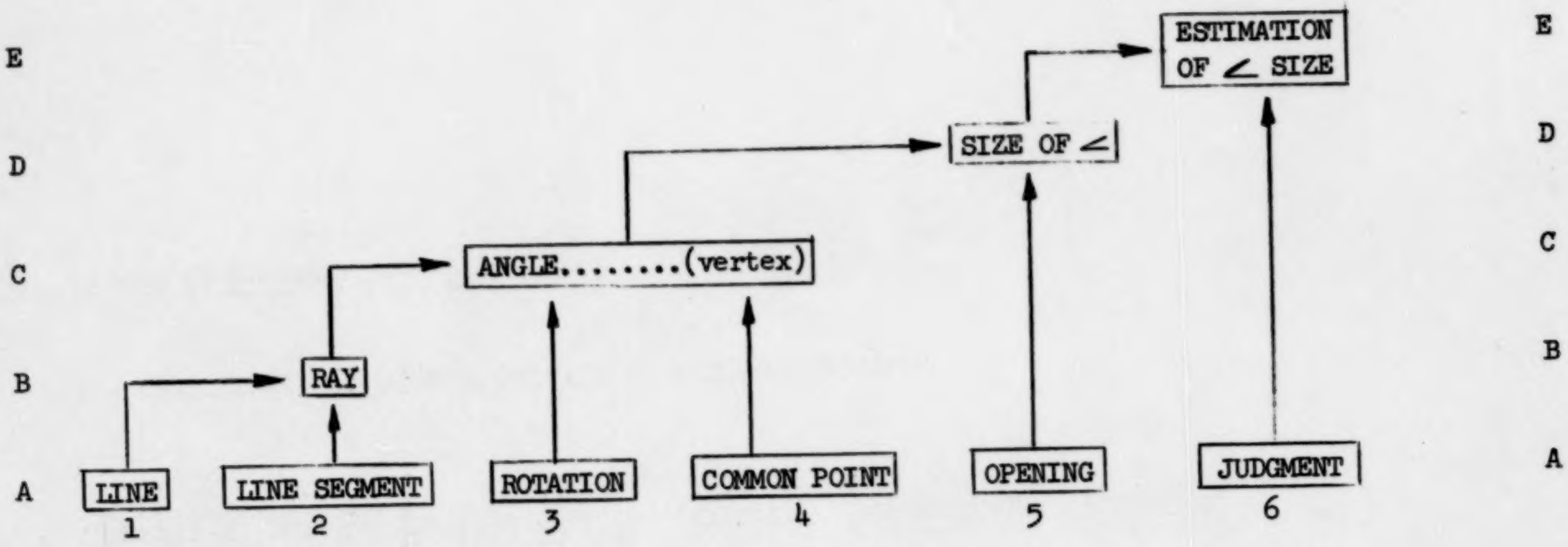
F. Provide for practice on various cueing systems on level of complexity generally found in tests at the high school level.

    1. Relationship of parts to whole.

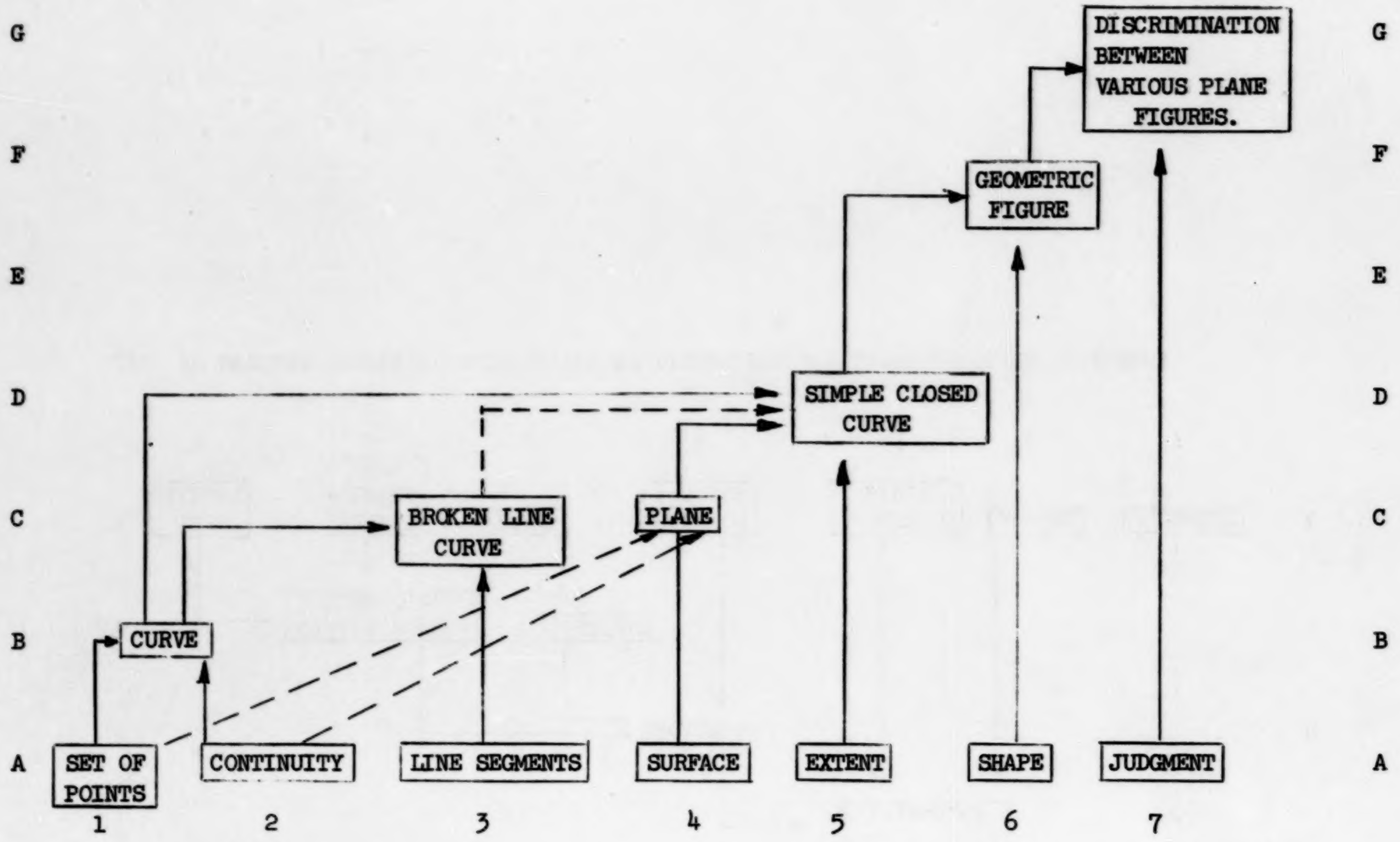
    2. Personal orientation to object image.


G
F
E
D
C
B
A
LOCATION
1
DOT
2
QUANTITY
3
SEQUENCE
4
EXTENT
5
MEASUREMENT
6
ESTIMATION
7
POINT
SET
LINE
LINE SEGMENT
LENGTH
ESTIMATION OF LENGTH
G
F
E
D
C
B
A
increasing levels of complexity
order of presentation

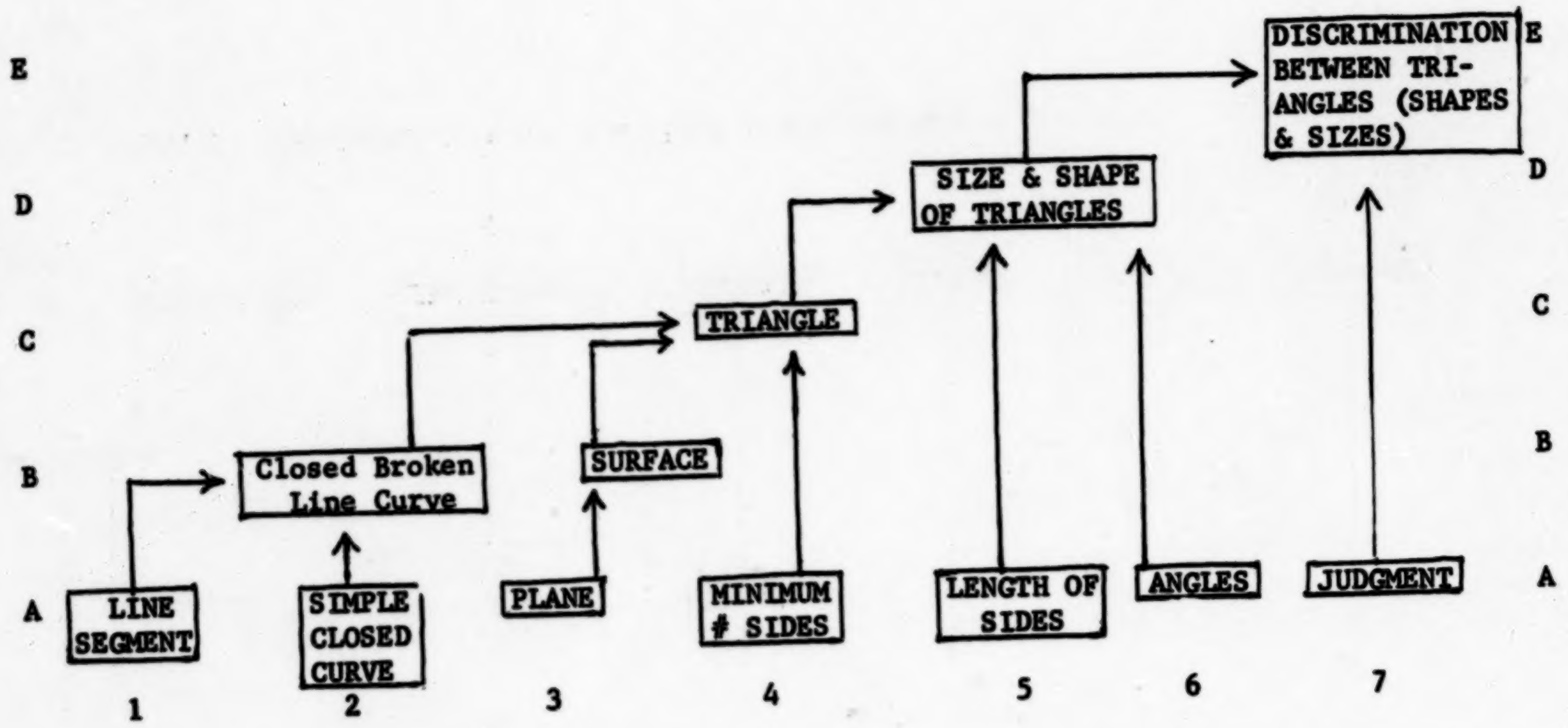
I. OBJECTIVE: TO DEVELOP ABILITY TO ESTIMATE LENGTHS.


E
D
C
B
A
LINE
1
LINE SEGMENT
2
RAY
ROTATION
3
COMMON POINT
4
ANGLE........(vertex)
OPENING
5
SIZE OF ∠
JUDGMENT
6
ESTIMATION OF ∠ SIZE
E
D
C
B
A

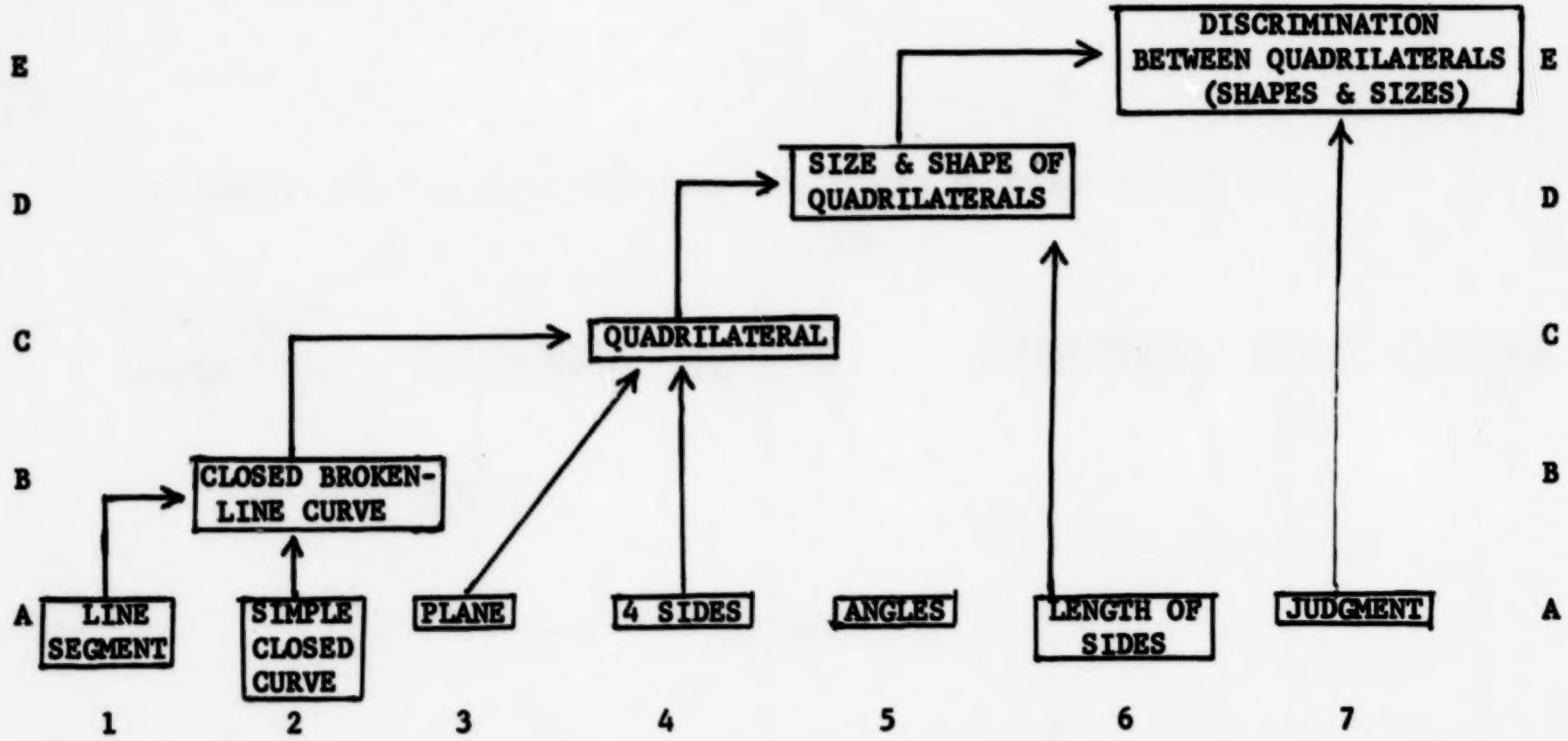
II. TO DEVELOP ABILITY TO ESTIMATE ∠ SIZE (DISCRIMINATE BETWEEN ANGLES USING RT. ∠ AS REFERENCE)


G
F
E
D
C
B
A
DISCRIMINATION BETWEEN VARIOUS PLANE FIGURES.
GEOMETRIC FIGURE
SIMPLE CLOSED CURVE
BROKEN LINE CURVE
PLANE
CURVE
SET OF POINTS
CONTINUITY
LINE SEGMENTS
SURFACE
EXTENT
SHAPE
JUDGMENT
1
2
3
4
5
6
7

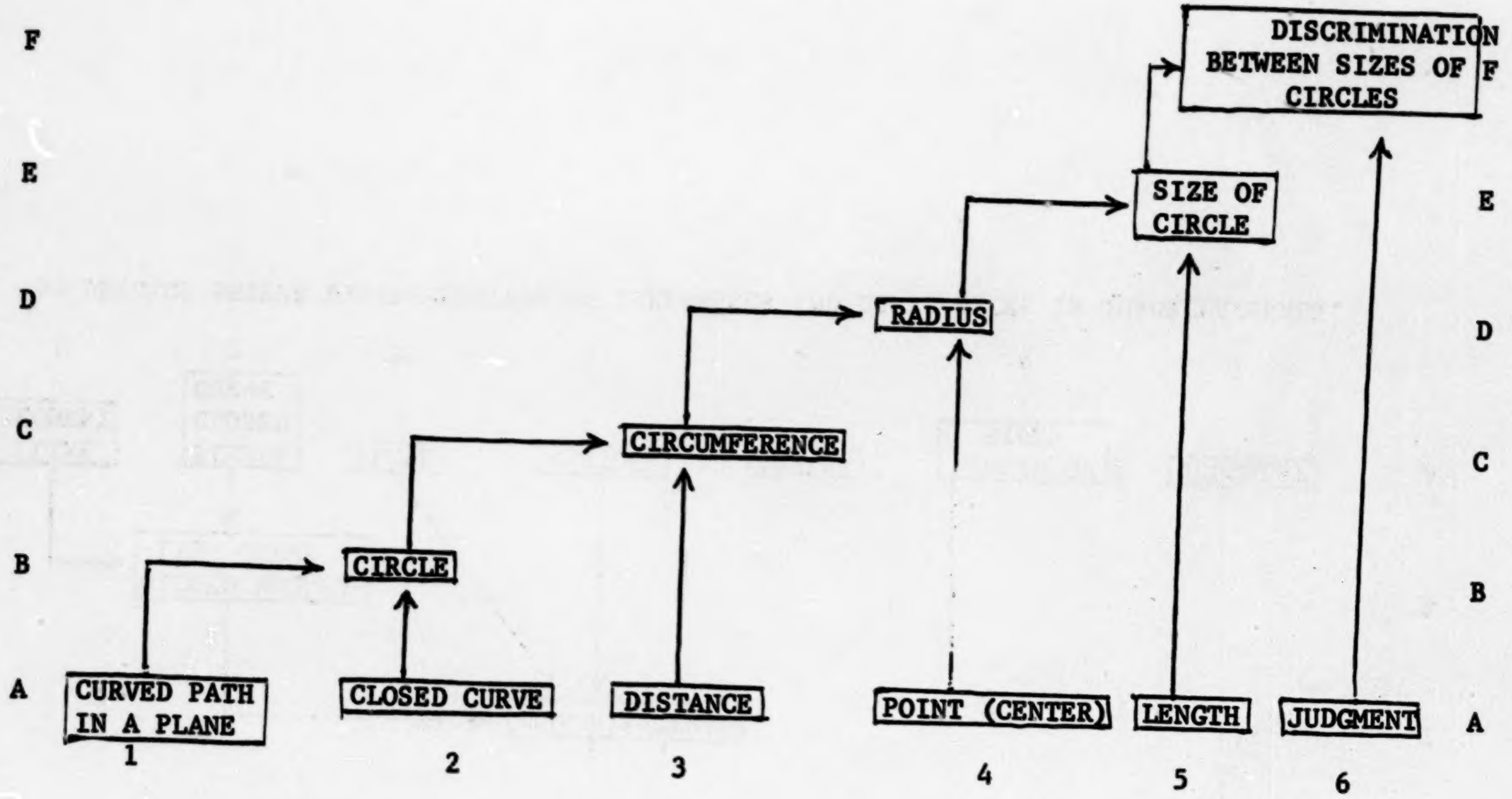
III. TO LEARN TO DISCRIMINATE BETWEEN VARIOUS PLANE FIGURES.


E
D
C
B
A
DISCRIMINATION BETWEEN TRIANGLES (SHAPES & SIZES)
SIZE & SHAPE OF TRIANGLES
TRIANGLE
Closed Broken Line Curve
SURFACE
LINE SEGMENT
SIMPLE CLOSED CURVE
PLANE
MINIMUM # SIDES
LENGTH OF SIDES
ANGLES
JUDGMENT
1
2
3
4
5
6
7

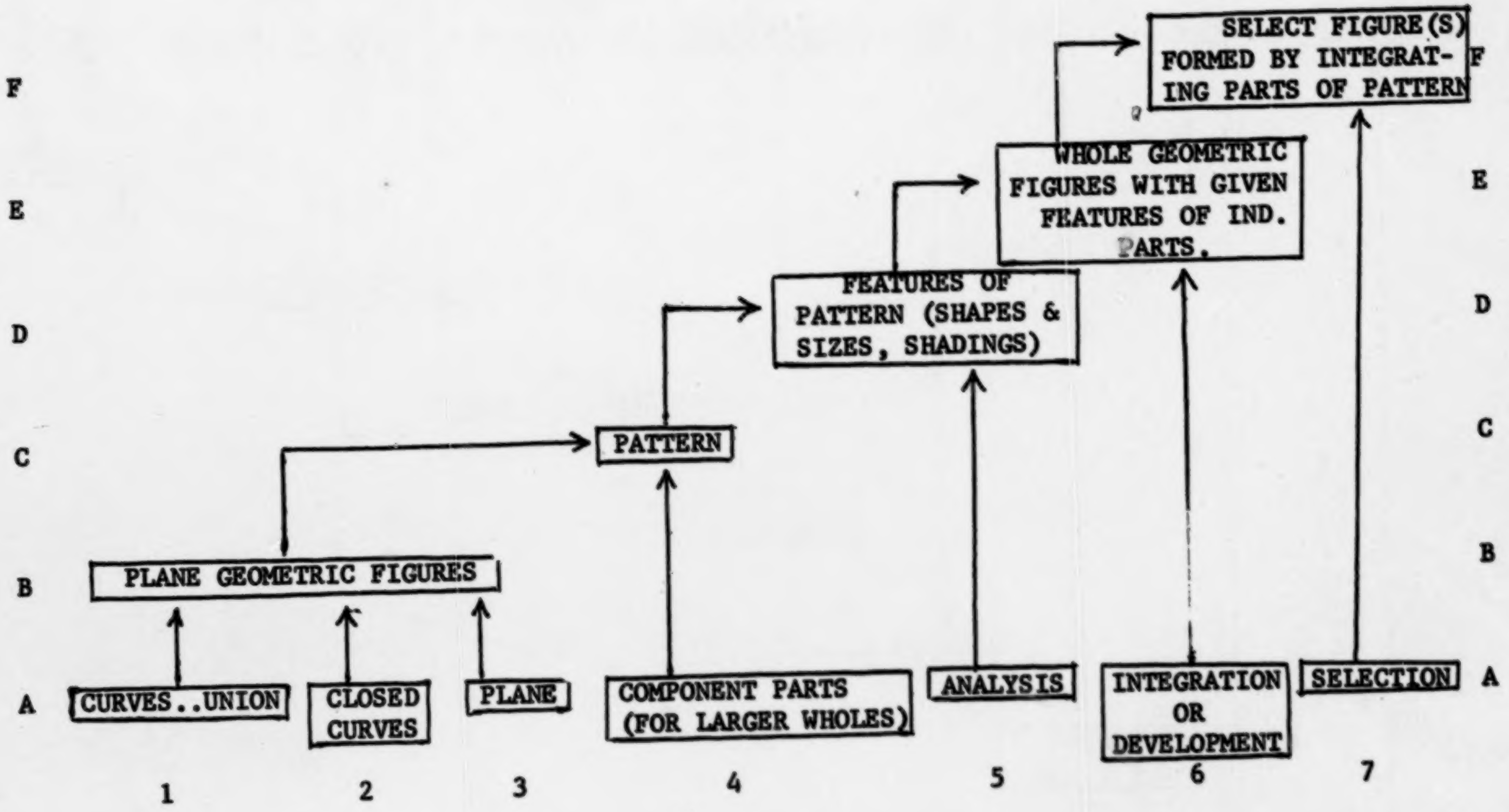
IV. TO DEVELOP BETTER DISCRIMINATION OF LIKENESSES AND DIFFERENCES IN TRIANGLES.


E
D
C
B
A
DISCRIMINATION BETWEEN QUADRILATERALS (SHAPES & SIZES)
SIZE & SHAPE OF QUADRILATERALS
QUADRILATERAL
CLOSED BROKEN-LINE CURVE
LINE SEGMENT
SIMPLE CLOSED CURVE
PLANE
4 SIDES
ANGLES
LENGTH OF SIDES
JUDGMENT
1
2
3
4
5
6
7
E
D
C
B
A

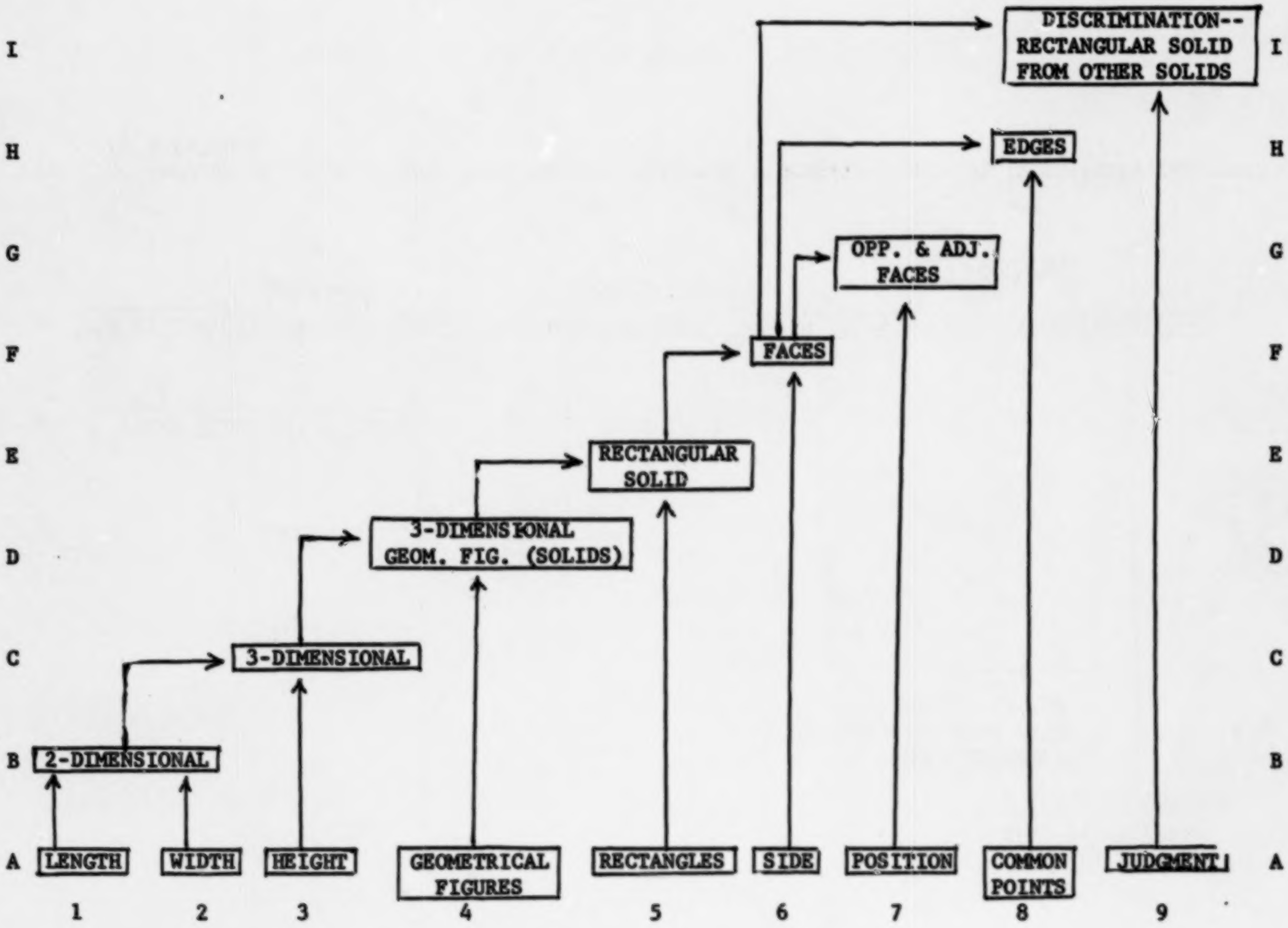
V. TO DEVELOP BETTER DISCRIMINATION OF LIKENESSES AND DIFFERENCES IN QUADRILATERALS.


F
E
D
C
B
A
CURVED PATH IN A PLANE
CLOSED CURVE
DISTANCE
POINT (CENTER)
LENGTH
JUDGMENT
1
2
3
4
5
6
CIRCLE
CIRCUMFERENCE
RADIUS
SIZE OF CIRCLE
DISCRIMINATION BETWEEN SIZES OF CIRCLES
F
E
D
C
B
A

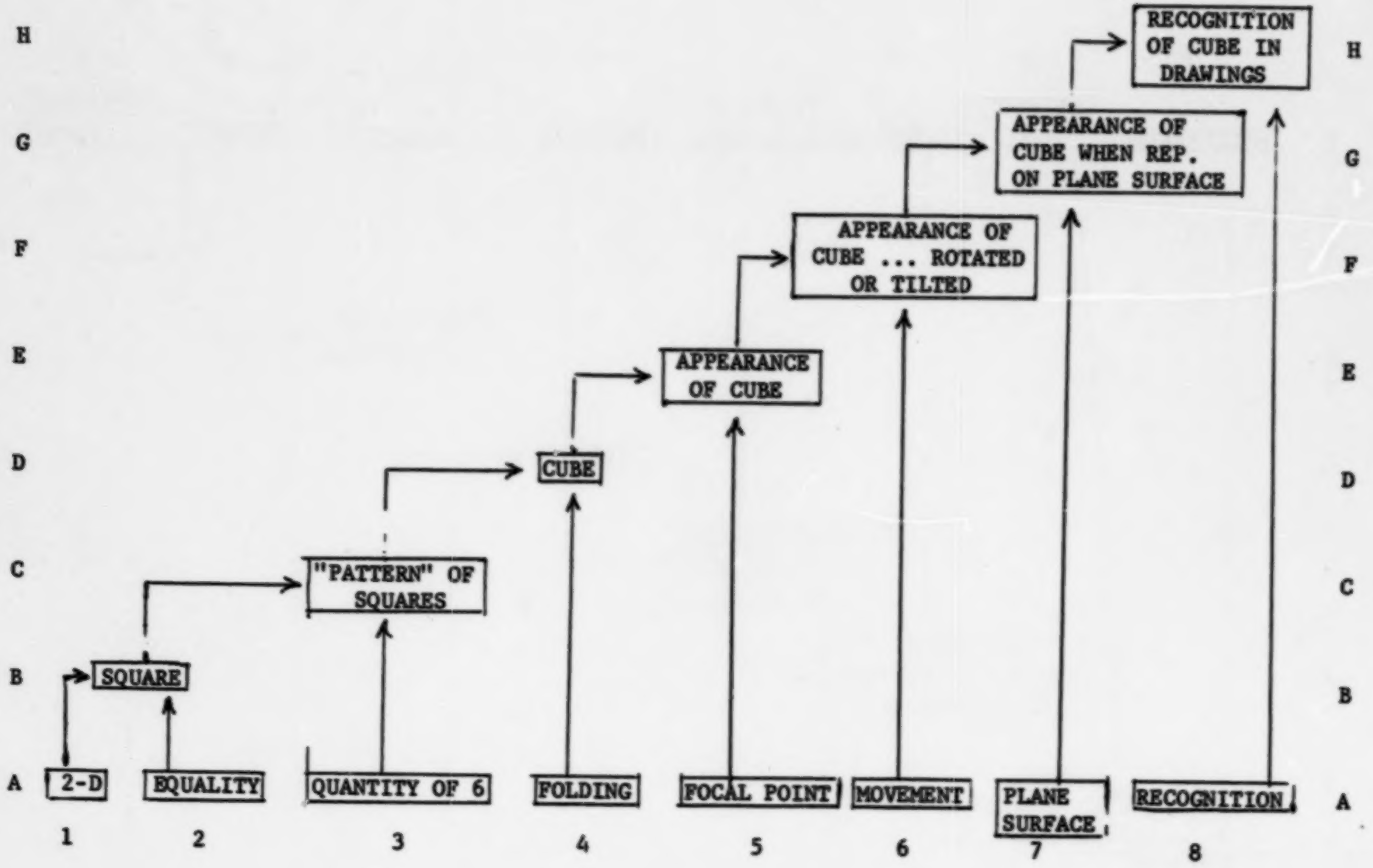
VI. TO DEVELOP ABILITY TO DISCRIMINATE DIFFERENCES IN SIZES OF CIRCLES.


F
E
D
C
B
A
SELECT FIGURE(S) FORMED BY INTEGRATING PARTS OF PATTERN
WHOLE GEOMETRIC FIGURES WITH GIVEN FEATURES OF IND. PARTS.
FEATURES OF PATTERN (SHAPES & SIZES, SHADINGS)
PATTERN
PLANE GEOMETRIC FIGURES
CURVES..UNION
CLOSED CURVES
PLANE
COMPONENT PARTS (FOR LARGER WHOLES)
ANALYSIS
INTEGRATION OR DEVELOPMENT
SELECTION
1
2
3
4
5
6
7
F
E
D
C
B
A

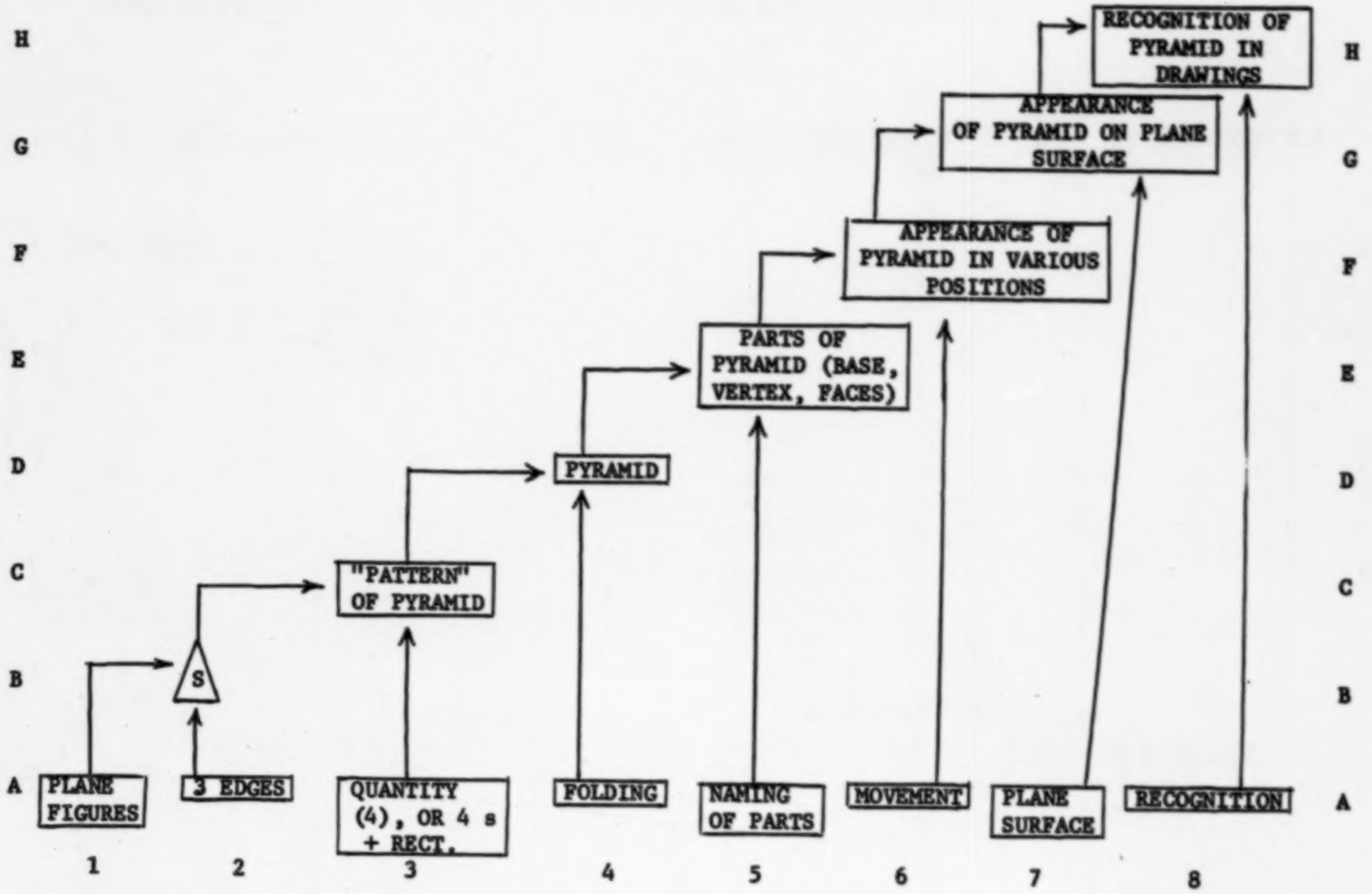
VII. TO PROVIDE PRACTICE IN SELECTING CORRECT GEOMETRIC FIGURE(S) FORMED BY INTEGRATING THE PARTS OF A PATTERN.


I
H
G
F
E
D
C
B
A
DISCRIMINATION--
RECTANGULAR SOLID
FROM OTHER SOLIDS
EDGES
OPP. & ADJ.
FACES
FACES
RECTANGULAR
SOLID
3-DIMENSIONAL
GEOM. FIG. (SOLIDS)
3-DIMENSIONAL
2-DIMENSIONAL
LENGTH
WIDTH
HEIGHT
GEOMETRICAL
FIGURES
RECTANGLES
SIDE
POSITION
COMMON
POINTS
JUDGMENT
1
2
3
4
5
6
7
8
9

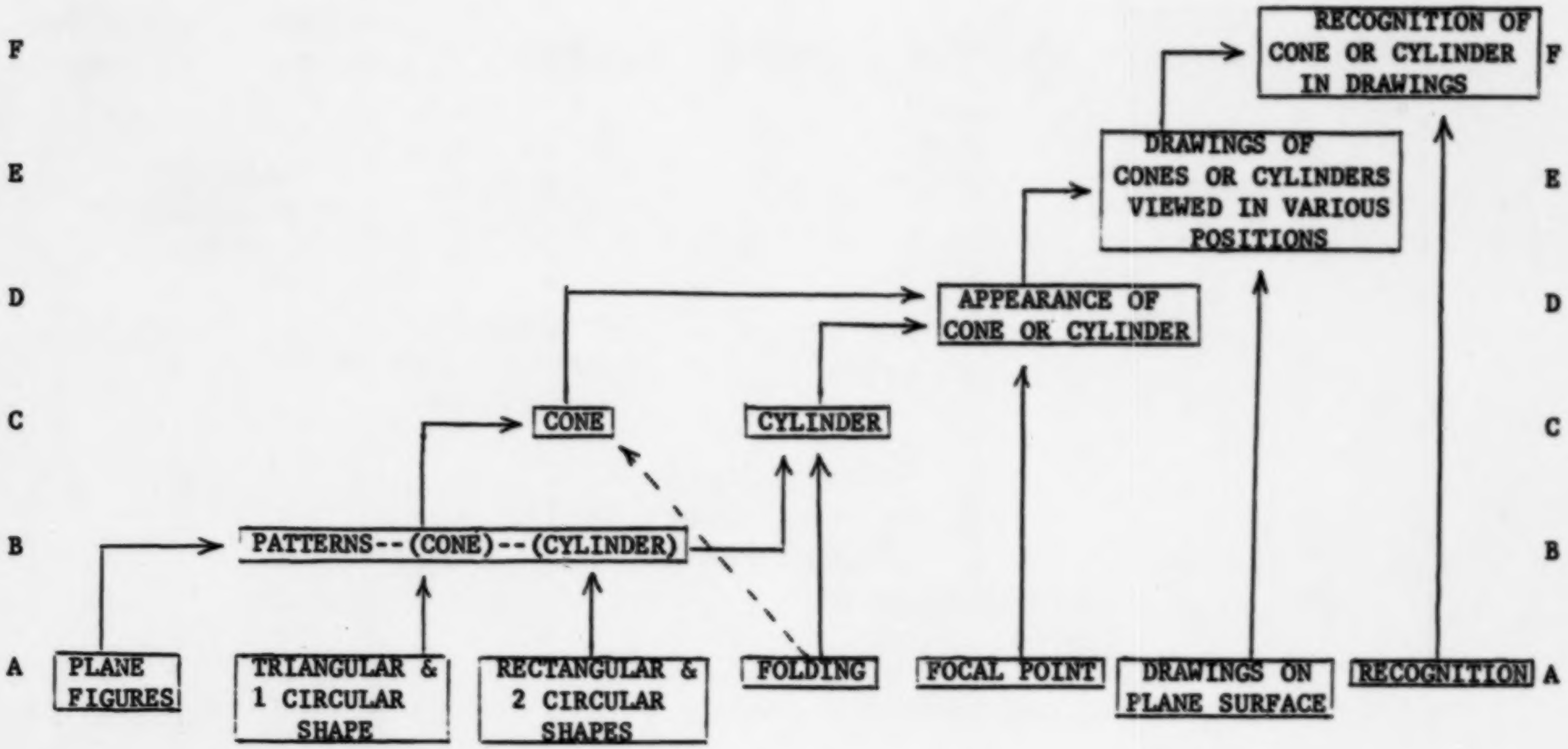
VIII. INTRODUCTION TO SOLIDS: LEARN TO DISCRIMINATE RECTANGULAR SOLIDS.


H
G
F
E
D
C
B
A
RECOGNITION OF CUBE IN DRAWINGS
APPEARANCE OF CUBE WHEN REP. ON PLANE SURFACE
APPEARANCE OF CUBE ... ROTATED OR TILTED
APPEARANCE OF CUBE
CUBE
"PATTERN" OF SQUARES
SQUARE
2-D
EQUALITY
QUANTITY OF 6
FOLDING
FOCAL POINT
MOVEMENT
PLANE SURFACE
RECOGNITION
1
2
3
4
5
6
7
8

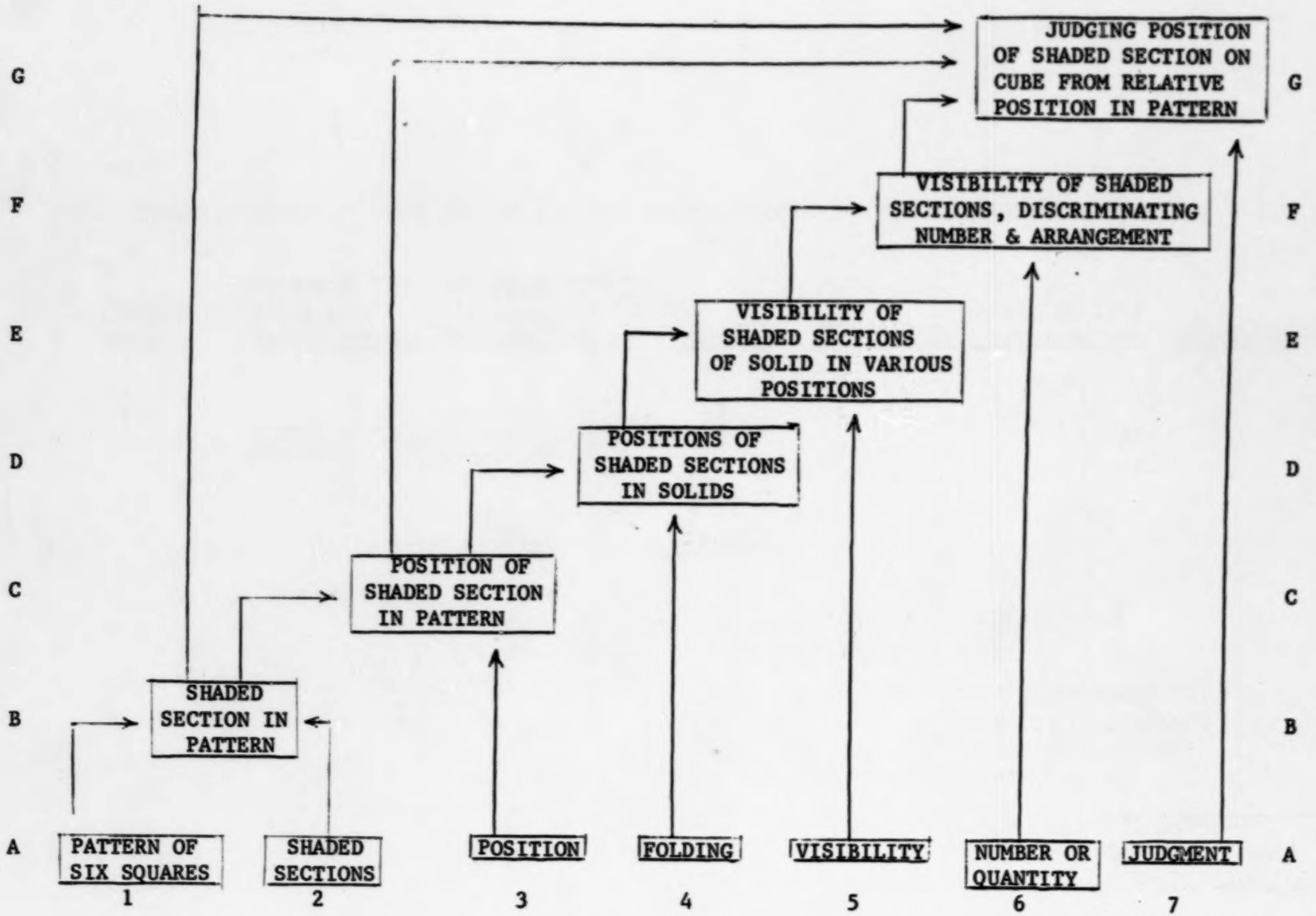
IX. LEARN TO RECOGNIZE CUBE AS REPRESENTED ON A PLANE SURFACE DRAWING.


H
G
F
E
D
C
B
A
RECOGNITION OF PYRAMID IN DRAWINGS
APPEARANCE OF PYRAMID ON PLANE SURFACE
APPEARANCE OF PYRAMID IN VARIOUS POSITIONS
PARTS OF PYRAMID (BASE, VERTEX, FACES)
PYRAMID
"PATTERN" OF PYRAMID
S
PLANE FIGURES
3 EDGES
QUANTITY (4), OR 4 s + RECT.
FOLDING
NAMING OF PARTS
MOVEMENT
PLANE SURFACE
RECOGNITION
1
2
3
4
5
6
7
8

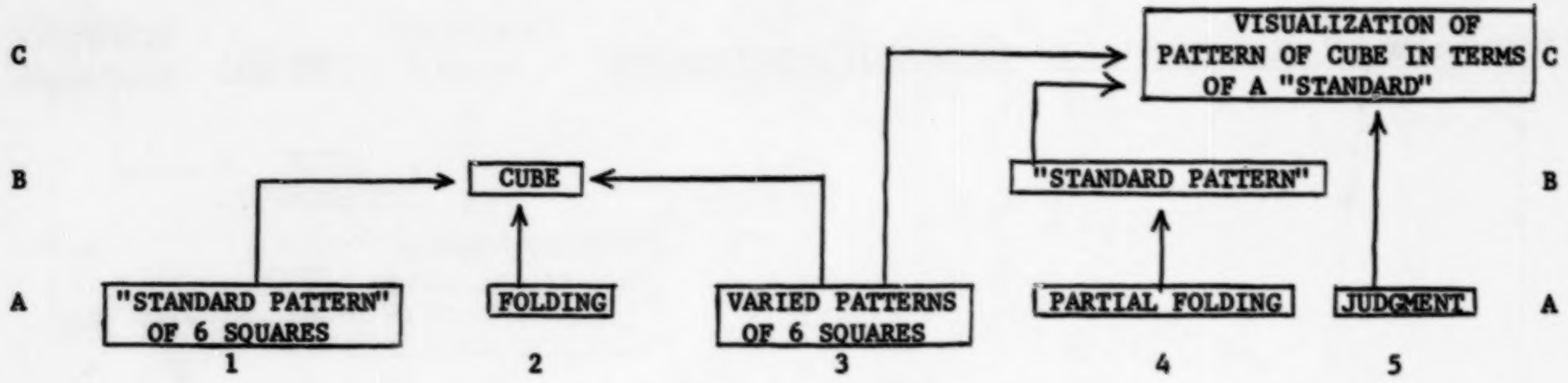
X. LEARN TO RECOGNIZE PYRAMID AS REPRESENTED ON A PLANE SURFACE DRAWING.


F
E
D
C
B
A
RECOGNITION OF CONE OR CYLINDER IN DRAWINGS
DRAWINGS OF CONES OR CYLINDERS VIEWED IN VARIOUS POSITIONS
APPEARANCE OF CONE OR CYLINDER
CONE
CYLINDER
PATTERNS--(CONE)--(CYLINDER)
PLANE FIGURES
TRIANGULAR & 1 CIRCULAR SHAPE
RECTANGULAR & 2 CIRCULAR SHAPES
FOLDING
FOCAL POINT
DRAWINGS ON PLANE SURFACE
RECOGNITION
F
E
D
C
B
A

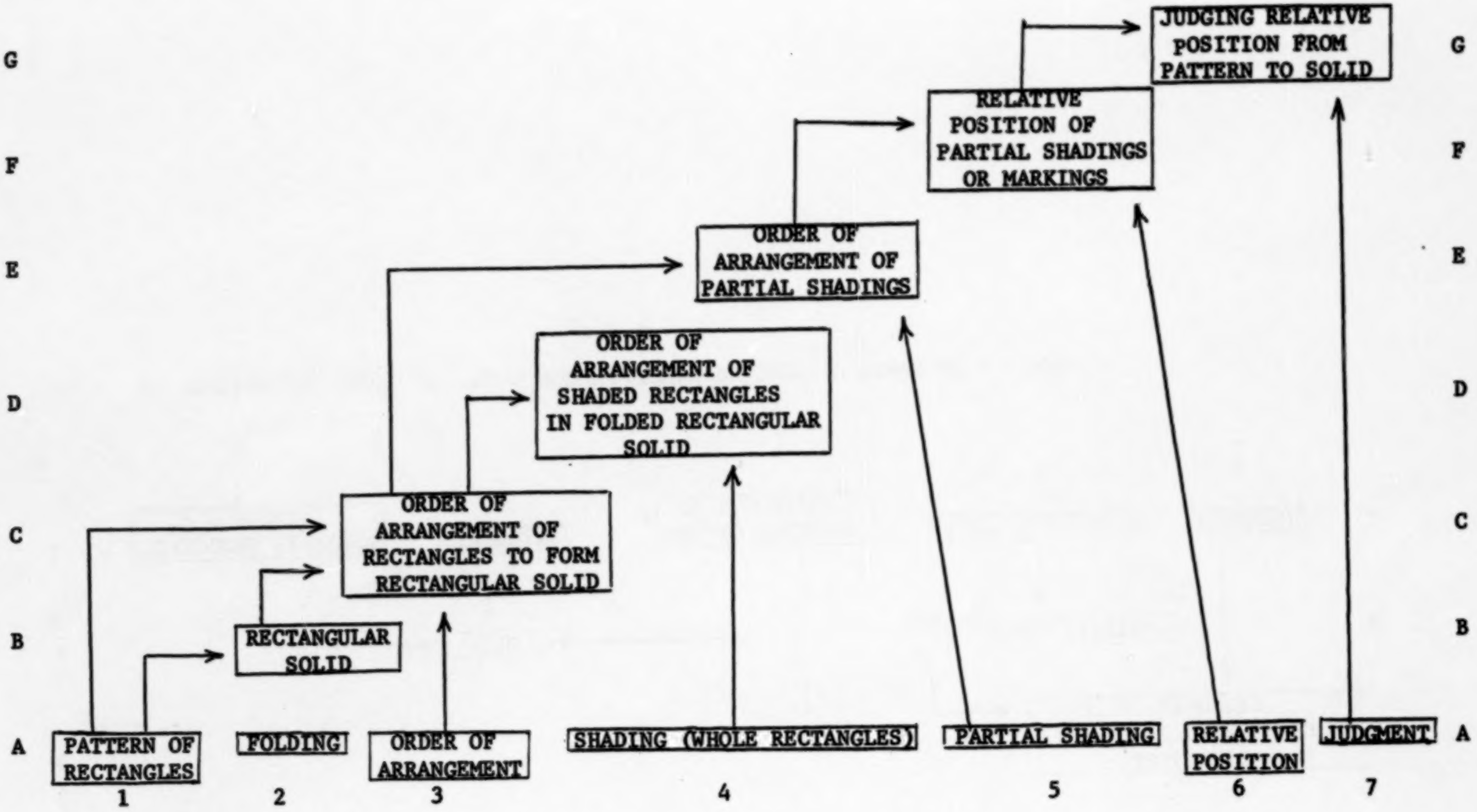
XI. LEARN TO RECOGNIZE CONE AND/OR CYLINDER AS REPRESENTED ON A PLANE SURFACE DRAWING.


G
F
E
D
C
B
A
JUDGING POSITION OF SHADED SECTION ON CUBE FROM RELATIVE POSITION IN PATTERN
VISIBILITY OF SHADED SECTIONS, DISCRIMINATING NUMBER & ARRANGEMENT
VISIBILITY OF SHADED SECTIONS OF SOLID IN VARIOUS POSITIONS
POSITIONS OF SHADED SECTIONS IN SOLIDS
POSITION OF SHADED SECTION IN PATTERN
SHADED SECTION IN PATTERN
PATTERN OF SIX SQUARES
1
SHADED SECTIONS
2
POSITION
3
FOLDING
4
VISIBILITY
5
NUMBER OR QUANTITY
6
JUDGMENT
7

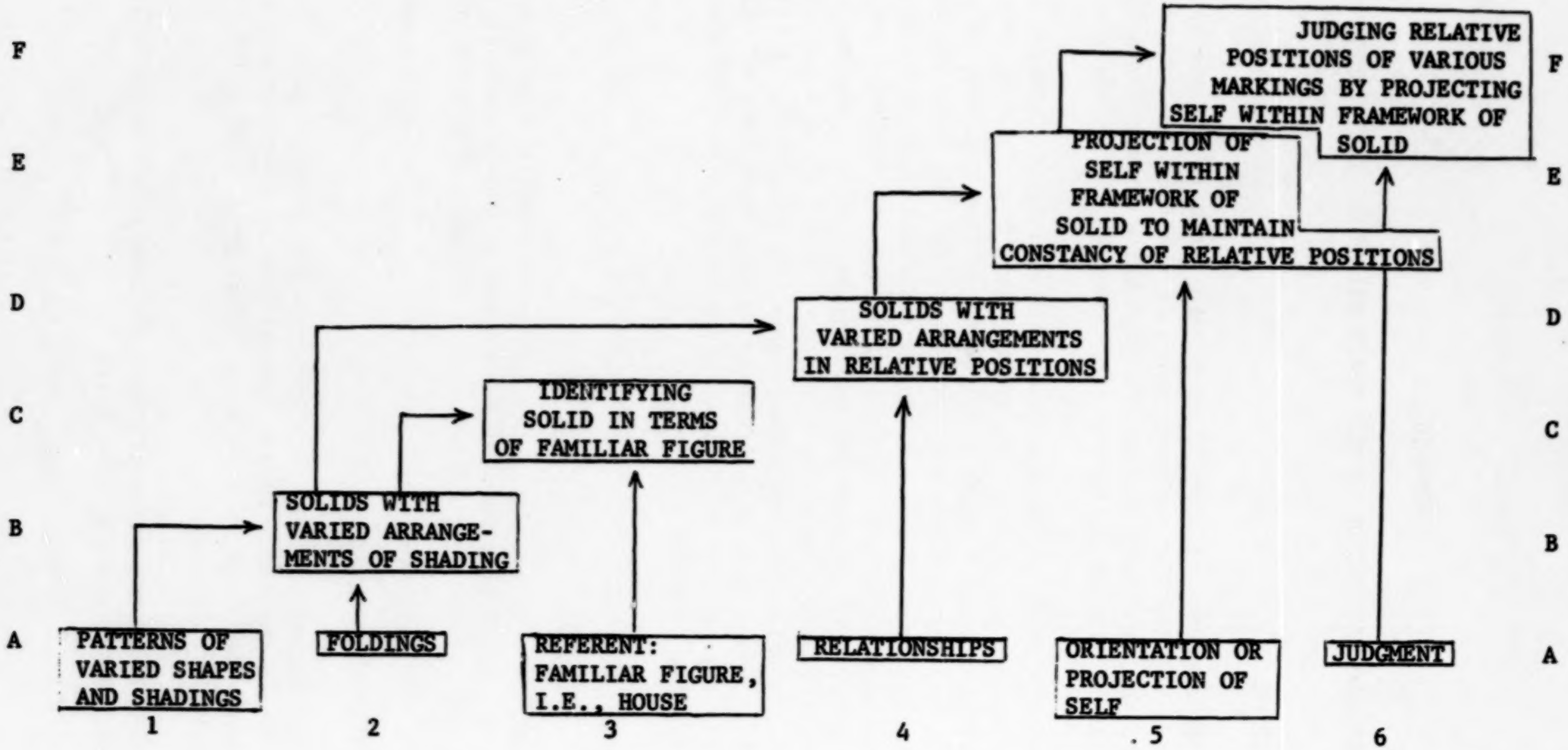
XII. TO PROVIDE PRACTICE IN JUDGING POSITION OF SHADED SECTION ON CUBE FROM RELATIVE POSITION IN PATTERN.


C
B
A
"STANDARD PATTERN" OF 6 SQUARES
1
CUBE
FOLDING
2
VARIED PATTERNS OF 6 SQUARES
3
"STANDARD PATTERN"
PARTIAL FOLDING
4
VISUALIZATION OF PATTERN OF CUBE IN TERMS OF A "STANDARD"
JUDGMENT
5
C
B
A

XIII. TO DEVELOP THE HABIT OF VISUALIZING PATTERN OF CUBE IN TERMS OF A STANDARD.


G
F
E
D
C
B
A
JUDGING RELATIVE POSITION FROM PATTERN TO SOLID
RELATIVE POSITION OF PARTIAL SHADINGS OR MARKINGS
ORDER OF ARRANGEMENT OF PARTIAL SHADINGS
ORDER OF ARRANGEMENT OF SHADED RECTANGLES IN FOLDED RECTANGULAR SOLID
ORDER OF ARRANGEMENT OF RECTANGLES TO FORM RECTANGULAR SOLID
RECTANGULAR SOLID
PATTERN OF RECTANGLES
FOLDING
ORDER OF ARRANGEMENT
SHADING (WHOLE RECTANGLES)
PARTIAL SHADING
RELATIVE POSITION
JUDGMENT
1
2
3
4
5
6
7

XIV. TO PROVIDE PRACTICE IN JUDGING POSITION OF SHADED SECTION ON RECTANGULAR SOLID FROM RELATIVE POSITION WITHIN PATTERN.



F
E
D
C
B
A
JUDGING RELATIVE POSITIONS OF VARIOUS MARKINGS BY PROJECTING SELF WITHIN FRAMEWORK OF SOLID
PROJECTION OF SELF WITHIN FRAMEWORK OF SOLID TO MAINTAIN CONSTANCY OF RELATIVE POSITIONS
SOLIDS WITH VARIED ARRANGEMENTS IN RELATIVE POSITIONS
IDENTIFYING SOLID IN TERMS OF FAMILIAR FIGURE
SOLIDS WITH VARIED ARRANGE-MENTS OF SHADING
PATTERNS OF VARIED SHAPES AND SHADINGS
FOLDINGS
REFERENT: FAMILIAR FIGURE, I.E., HOUSE
RELATIONSHIPS
ORIENTATION OR PROJECTION OF SELF
JUDGMENT
1
2
3
4
5
6

XV. TO UTILIZE TECHNIQUE OF PROJECTING SELF WITHIN FRAMEWORK OF SOLID TO MAINTAIN CONSTANT VIEW OF RELATIVE POSITIONS.

APPENDIX B

DIRECTIONS FOR ADMINISTRATION OF THE SPACE RELATIONS TEST (PRETEST)

STEP 1. INTRODUCTION

Try to put the students at ease by explaining briefly why the test is being administered. Stress the personal value of the test for each student, so that the pupils will not only accept the test but also put forth their best efforts. Read the following introductory remarks to the class:

This is part of a study conducted by The University of Michigan to see how well 8th-graders can do on tests in space relations. In this test you will be shown various patterns which could be foled into figures. You are to decide which of the various figures shown could be made from the patterns. The ability to do this type of activity is considered important in many kinds of work such as machine-shop work, carpentry, mechanical drawing, dressmaking, operating a motor vehicle (automobile, airplane, etc.), engineering, art, construction work, and many other jobs. Tests of this type are also used to help decide which students are permitted to enter various kinds of training after high school. So you see, being able to do well in a test of this type may help you to get the job you want or help you to get into the type of training you want after high school.

STEP 2. PENCILS

Each pupil should have available two sharpened pencils with erasers. Perhaps each teacher could have available an extra supply of pencils to be made available to the student should he break his pencil during the test.

STEP 3. ANSWER SHEETS

Distribute the answer sheets; then instruct the group to print the requested information at the top. To avoid a barrage of questions, you may wish to dictate what should be on each blank line, pausing to allow time for the printing of information.

Special instructions are necessary when the group has never before used IBM answer sheets. Say something like this:

"In this test all the questions are in special booklets. You will write nothing in these booklets. You will mark your answers on the special answer sheets. Notice the many sets of small spaces on these answer sheets. (Give a more detailed description.) You are to show your choice of answer by marking a clear, heavy black mark in the proper space, like this (draw an example on the board and mark it). I want to emphasize now that you are to mark clearly, make your marks black, and you are to erase any marks other than your answers. You are to make your marks between the lines as I have shown you."

STEP 4. BOOKLETS

Before distributing the booklets for the test, make the following announcement so that it is clearly understood:

"Do not open these booklets until I tell you to do so."

Make certain that all students heed this instruction. Also stress again that absolutely no marks are to be made in the booklets; all marks are to go on the answer sheets.

STEP 5. READING DIRECTIONS

Directions for the test are in the test booklet. They are to be read aloud by you. Tell the students to read the directions to themselves as you read aloud. Read each example aloud in its entirety as printed.

After reading the directions, ask if there are any questions. In order to keep standard testing conditions, when answering questions, reread appropriate sections of the directions or examples. Make certain that each student understands what he is to do before beginning the test. Quietly given help on the first item or two is permitted to correct procedure. However, you should only clarify directions and procedure; you should not indicate the correct answer.

STEP 6. TIMING

After directions have been read and questions answered, in a firm voice say, "Begin," and start timing. The exact time allowed is 30 minutes. Make sure all students stop promptly when you say, "Stop."

STEP 7. COLLECTING MATERIALS

Collect the answer sheets, then the test booklets. Be sure every item is collected.

STEP 8. TESTING LOG

Immediately following the testing session, indicate deviations from regular procedures, names of those tested for whom any unusual factors seemed to be operating, and further information which might help in interpreting test results.

DIRECTIONS FOR ADMINISTRATION OF THE SPACE RELATIONS TEST (POSTTEST)

The posttest directions were the same as the pretest directions (on preceding pages) with one addition. The introductory remarks were prefaced with the following paragraph.

"Several weeks ago you took part in a study conducted by The University of Michigan to see how well 8th-graders could do on tests in space relations. Many of you showed that you could do very well on a test of this type. In giving tests of this kind it is often found that students can do better simply by taking another test of the same type. Also, you may have learned more about this material on your own since the last test. Today you will have another chance to show what you can do. *It is important that you do your very best*.

APPENDIX C

INSTRUCTIONS FOR GEOMETRIC INVENTORY (PRETEST)

STEP 1. INTRODUCTION (Read the following remarks:)

"This (Hold up a copy of the Geometrical Inventory.) is part of a study to find out how much 8th-graders already know about certain ideas related to simple figures which are often included in various courses of arithmetic or mathmetics. Some of you may know more of the answers than others. If you think you know an answer, write it on your answer sheet (not on the question form). There will be no penalty for guessing. Read each of the Read each of the questions carefully and then write your answer on the answer sheet. If you cannot even guess what the answer may be, then leave that space on your answer sheet blank and go on to the next question."

STEP 2. Distribute answer sheet No. 1 only to the class. Hve each student write his name in the place indicated. Since this is more of a program form than a test form, the sequential nature of the items may mean that later items "give away' answers for some of the earlier items, hence the need to collect answer sheet No. 1 before beginning on No. 2. Continue as follows:

"Place your name and today's date at the top of the page. After you finish page 1, turn in this answer sheet and pick up answer sheet No. 2. Be sure to write your name on answer sheet No. 2 also. Read each of the questions carefully and answer as many as you can. When you are finished, return your second answer sheet and the question form to your teacher."

STEP 3. Distribute Geometric Inventory forms and tell pupils to begin work. There is no time limit but if students spend too much time on some items encourage them to proceed. They should be able to complete this form in 15 to 20 minutes.

STEP 4. Place all materials (answer sheets and question forms) in this envelope and return to the principal's office.

(Note To Teachers: Try to allay any unnecessary anxiety which may occur after completion of the form for those students doing poorly. They are not expected to do very well on the form. You might tell them that this is to find out what should and what should not be included in a course of mathematics—not to grade them.)

INSTRUCTIONS FOR GEOMETRIC INVENTORY (POSTTEST)

STEP 1. INTRDOUCTION (Read the following remarks:)

"Several weeks ago you took part in a survey to help The University of Michigan find out how much you already knew about certain ideas related to simple figures which are often included in various courses of arithmetic or mathematics. Many of you showed that you already knew quite a few things about this material. In giving tests of this kind it is often found that students can do better simply by taking another test of the same type. Also, you may have learned more about this material on your own since the last test.

Today you will have another chance to show what you can do. It is important that you do your best. If you think you know an answer, write it on your answer sheet (not on the question form). There will be no penalty for guessing. Read each of the questions carefully and then write your answer on the answer sheet. If you feel that you cannot even guess what the answer may be, then leave that space on your answer sheet blank and go on to the next question."

STEP 2. Distribute answer sheet No. 1 only to the class. Have each student write his name in the place indicated. Since this is more of a program form than a test form, the sequential nature of the items may mean that later items "give away" answers for some of the earlier items, hence the need for collecting answer sheet No. 1 before beginning on No. 2. Continue as follows:

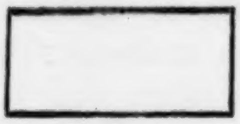
"Place your name and today's date at the top of the answer sheet. After you finish page 1, turn in this answer sheet and pick up answer sheet No. 2. Be sure to write your name on answer sheet No. 2 also. Read each of the questions carefully and answer as many as you can. When you are finished, return your second answer sheet and the question form to your teacher."

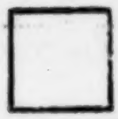
STEP 3. Distribute Geometric Inventory forms and tell students to begin work. There is no time limit but if students spend too much time on some items, encourage him to proceed.

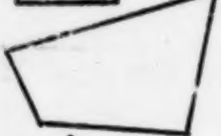
APPENDIX D

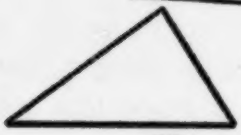
GEOMETRIC INVENTORY (SAMPLE COPY)

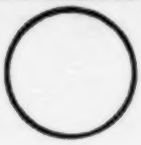
I. Identify each of the following by writing its name on the answer sheet opposite the correct number.

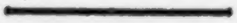
1.


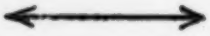
2.


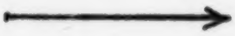
3.


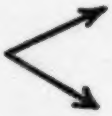
4.


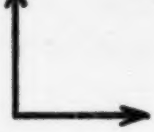
5.


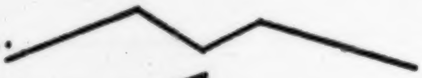
6.


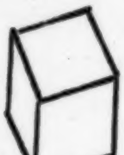
7.


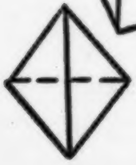
8.


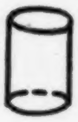
9.


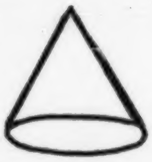
10.


11.


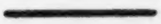
12.


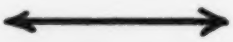
13.


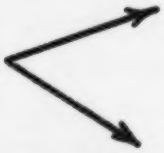
14.


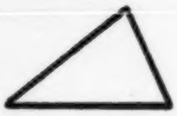
15.


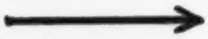
II. Write the correct symbol for each of the following:

16.


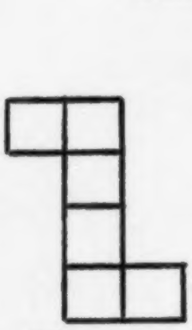
17.


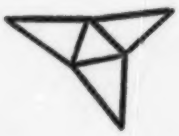
18.


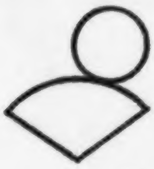
19.


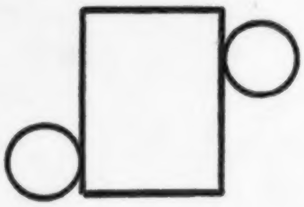
20.


III. Below are drawings of patterns each of which could be folded to form a solid. Write the name of the solid which can be formed from each pattern:


21.


22.


23.


24.

IV. In the following supply the word (or words) as indicated by the blanks: (Write your answer on the answer sheet provided.)

25. Collections such as flock, team, group, and class can be called a ______________________.

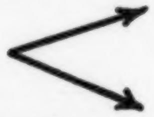
26. These two drawings A_____B are of the same length; we say that they
C_____D
are equal or ______________________.

27. A plane is a ____________________ dimensional space.

28. A solid is a ____________________ dimensional space.

29. For two plane geometric figures to be alike they must have the same ______________, ______________, and markings or shadings.

30. We may use a small dot to represent the location of a place or ________.

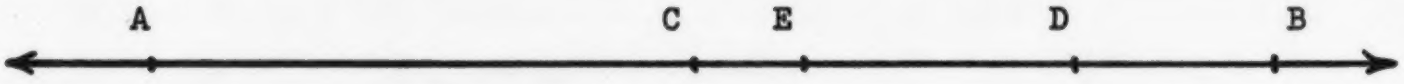
31. In this drawing  the point B is called the ____________.

32. To say that two triangles are congruent means that they are ___________.

33. In order to form a cube all 6 of the parts of the pattern must be in the shape of a ____________________.

34. If we want to refer to the whole class of four-sided figures, we use the special name ____________________, which means "four-sided."

In the next 5 items refer to the drawing shown here and complete the statemtns with longer than, shorter than, or congruent with. Make your decision by just looking at the sections called for in the statements below.



35. AC is __________ __________ EB.

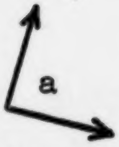
36. DB is __________ __________ CE.

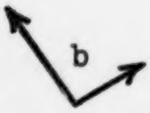
37. ED is __________ __________ DB.

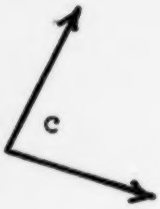
38. CD is __________ __________ AC.

39. CE is __________ __________ DB.

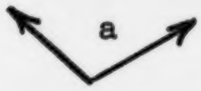
40. Which of the following angles is larger?

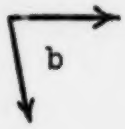


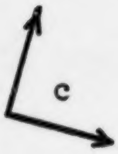


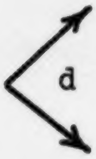


41. Which of these angles looks most like a right angle?

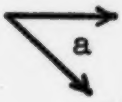


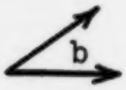


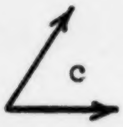


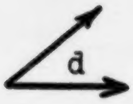


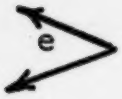
In the next 5 items refer to the drawings shown here and fill in the blanks with one of the following: larger than, smaller than, or congruent with.

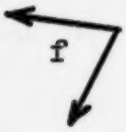












42. a is __________ __________ e.

43. d is __________ __________ b.

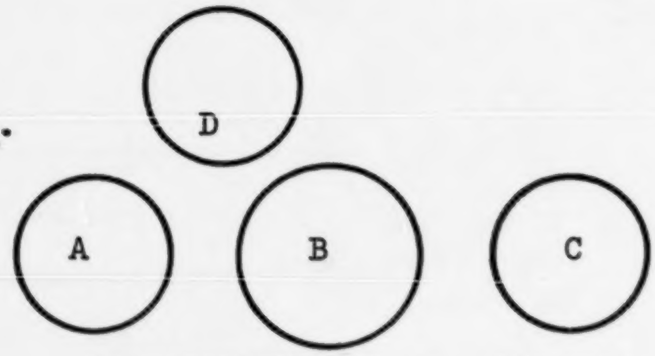
44. b is __________ __________ f.

45. c is __________ __________ e.

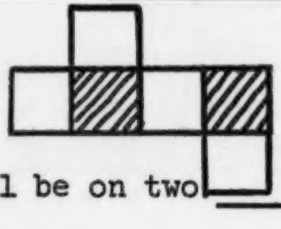
46. d is __________ __________ c.

47. The smallest circle is: __________.

48. The largest circle is: __________.

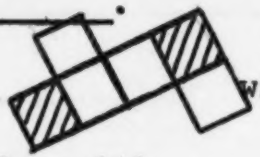


49. When this pattern is folded to form a solid, the shaded squares will be on two __________ sides.



50. When we look at a large sized cube at a distance, the most faces that we can see at any one time is ______________.

51. The folding of this pattern will put the shaded faces on two ______________ sides of the solid.



52. When 2 opposite faces of a cube are shaded, how many of the shaded faces may be visible when the cube is held so that 3 of its faces are visible?

____

53. Can the cube be held in such a position so that both adjacent shaded faces are hidden? ____________

54. When 3 shaded faces of a cube are arranged in "consecutive" order, can you hold the cube so that all three shaded faces are visible? __________

55. If the 3 shaded faces of a cube are adjacent, may all three be visible at the same time? ____________

56. When 3 shaded faces of a cube are adjacent to each other, they share a common ____________.

57. With the 3 shaded faces of a cube arranged "consecutively" and with 3 faces of the cube visible, one may be able to see as many as _______ shaded faces or as few as _______ shaded face(s) at any given time.

58. With the 3 shaded faces sharing a common vertes and with 3 faces of the cube visible, I may be able to see as many as __________ or as few as __________ shaded faces at any given time.

59. When we have two types of markings or shadings on a solid, we look for the relationship of the similar markings to _______ _________, to the second _________ and to the ________ itself.

60. For all markings of shadings on 3 faces of a cube or rectangle to be hidden when half of the solid is visible, they must be on ____________ sides.

APPENDIX E

DIRECTIONS FOR USING THE PROGRAMMED MATERIALS

STEP 1. Distribute Unit I of the Program. Read the introductory remarks found on the first part of Unit I. Go through the instruction page with the class.

STEP 2. Distribute the Panel Booklets and read through first page of instructions.

If further illustrations are necessary, feel free to provide them to clarify procedure. You may work the first few frames as a class in order to get the pupils to turn the pages properly. After completing No. 10, they are to return to page 1 as indicated. Some pupils may need help here. This is not a test so don't be too concerned about giving away answers. They are already provided for the pupil to see for himself. The student should not, of course, just copy the answer without even trying to construct his own. This defeats the whole purpose of the programmed learning. As indicated in the instructions, should a student find his response incorrect, he should draw a circle around his anser and then he may copy the correct answer after it.

STEP 3. General Management

You will be provided with each of the successive units in time for usage by the students. After a student completes a Unit, have him turn it in along with the answer sheet; then provide him with the next unit and answer sheet.

Materials are to be turned in at the end of class each day. All work is to be done in class. (Do not allow materials to go home.)

If a pupil has difficulty in grasping a certain point, the teacher may give some assistance in clarifying the point. If specific areas of the program seem to cause widespread confusion, make a note of this in your report.

Students may occasionally use an equivalent response but not identical to the response given. Use your judgment regarding the degree of equivalence if student asks whether his anser isn't correct.

The answer sheet provides a space for starting time and completion time. Try to have the students fill in the times. If a student happens to be only partly through when it is quitting time, have him turn in the materials and let him continue where he left oof the next day by placing that day's date in the left margin, (for example; he may have just finished number 14; have him write the next day's date in the margin preceding No. 15).

The improper use of panels may cause some problems for later users of the panels if the pupils remove the cut-out patterns and fail to replace them in the proper pockets. There will be no need to use the patterns until they rach panel 13, except perhaps to look at a pattern as you read the introductory page.

The boxes containing the solids will be brought in later. These will not be needed until the student begins with Unit V, (Item No. 235).

APPENDIX F

RAW DATA FOR EXPERIMENTAL AND CONTROL GROUPS

Data for the study were collected at East Junior High School, Ypsilanti, Michigan. The raw data are recorded according to the following code for the columns:

A	the pupil's code number
B	sex
C	pre-course score on *Space Relations Test*, Form A, of the *Differential Aptitude Tests*.
D	post-course score on *Space Relations Test*, Form B, of the *Differential Aptitude Tests*.
E	Score on *Lorge-Thorndike Intelligence Tests* (Total)
F	pre-course score on *Geometric Inventory*
G	post-course score on *Geometric Inventory*

For Experimental Group only:

H	total errors on program
I	total time (in minutes) on program
J	score on Reading Comprehension test (raw score) of the *Iowa Tests of Basic Skills*
K	score on Arithmetic concepts test (raw score) of the *Iowa Tests of Basic Skills*

I. Experimental Group Data

A	B	C	D	E	F	G	H	I	J	K
1	F	71	85	116	24	58	8	360	50	24
2	M	68	77	126	28	56	71	352	52	27
3	F	62	73	126	26	60	36	304	52	26
4	M	62	63	118	18	43	25	309	62	24
5	M	60	58	118	28	57	48	397	62	19
6	F	56	79	114	23	57	22	328	61	25
7	F	54	83	105	21	51	46	286	24	13
8	F	51	57	104	12	58	45	267	35	16
9	M	50	80	115	30	63	23	342	52	25
10	F	49	60	106	17	47	30	285	40	23
11	M	47	76	115	17	59	27	348	63	37
12	M	44	75	110	25	61	97	365	47	25
13	M	41	79	121	29	58	23	454	47	20
14	F	41	51	---	18	39	30	344	29	13
15	M	39	46	89	14	38	88	389	23	24
16	M	35	40	96	20	46	40	210	51	16
17	M	24	39	---	21	52	89	310	50	20
18	F	24	68	101	19	58	14	338	51	23
19	M	22	47	99	20	45	31	385	30	18
20	F	16	67	113	21	58	50	293	51	16
21	F	14	24	92	6	31	21	265	24	12
22	F	12	9	101	14	49	93	395	32	18
23	F	9	21	90	12	33	47	363	25	11
24	M	7	34	93	18	47	80	304	40	22
25	M	7	16	96	18	38	37	371	14	15

26	M	7	15	99	13	48	117	376	51	15
27	F	1	42	112	13	46	49	317	39	16

II. Control Group Data

A	B	C	D	E	F	G
28	F	67	68	115	23	24
29	M	69	73	131	23	32
30	F	66	85	116	18	17
31	M	66	76	113	31	35
32	M	60	81	117	31	39
33	F	57	71	116	22	27
34	F	55	67	116	31	29
35	F	50	43	101	15	29
36	M	58	65	120	28	32
37	F	44	47	118	31	25
38	M	47	31	116	29	31
39	M	40	52	110	22	23
40	M	42	45	---	23	25
41	F	42	34	---	23	28
42	M	38	32	100	16	15
43	M	36	16	94	16	18
44	M	20	29	90	12	22
45	F	23	26	100	20	23
46	M	14	3	105	29	31
47	F	16	8	107	27	27
48	F	15	17	94	9	3
49	F	14	16	103	17	24
50	F	18	12	96	23	19
51	M	10	24	90	--	22
52	M	0	0	91	13	--
53	M	5	15	94	6	9
54	F	2	24	109	13	21

REFERENCES

Anastasi, Anne. Psychological Testing. (2nd Ed.) New York: The Macmillan Co., 1961.

Arnoult, M. O. Transfer of pre-differentiation training in simple and multiple shape discrimination. Journal of Experimental Psychology, 1953, 45, 401-409.

Bennett, K., Seashore, H. G., and Wesman, A. G. Manual for the Differential Aptitude Tests. New York: The Psychological Corporation, 1959.

Blade, Mary F., and Watson, W. S. Increase in spatial visulization test scores during engineering study. Psychological Monograph, 1955, 69 (12), No. 397.

Blyth, J. W. The Hamilton College Experiment in Programmed Learning. Project report: supported by the Fund for the Advancement of Education of the Ford Foundation, 1961.

Brown, F. R. The Effect of an Experimental Course in Geometry on Ability to Visualize in Three Dimensions. Unpublished doctoral dissertation, University of Illinois, 1954.

Carpenter, F., and Haddan, E. E. Systematic Application of Psychology to Education. Macmillan Co., New York, 1964.

Crowder, N. A., and Zachert, Virginia. Use of the Autotutor at Keesler AFB to Train in Fundamentals of Electronics. Training Systems, Western Design and Electronics, Division of US Industries, Inc., Santa Barbara Airport, Goleta, California, 1960.

Della-Piana, G. An experimental evaluation of programmed learning. The Journal of Educational Research, 1962, 55, 476-484.

French, J.R.P., Jr. Experiments in field settings. In L. Festinger and D. Katz (eds.), Research Methods in the Behavioral Sciences. New York: The Dryden Press, 1953, 98-135.

French, J. W. The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monograph, No. 5, 1951.

Gagne, R. M., and Baker, Katherine E. Stimulus pre-differentiation as a factor in transfer of training. Journal of Experimental Psychology, 1950, 40, 439-451.

Galanter, E. Automatic Teaching: The State of the Art. New York: Wiley, 1959.

REFERENCES (Continued)

Graham, C. H. Visual perception. In S. S. Stevens (ed.), Handbook of Experimental Psychology. New York: Wiley, 1951, 868-920.

Green, E. J. The Learning Process and Programmed Instruction. New York: Holt, Rinehart, and Winston, Inc., 1962.

Green, E. J., Weiss, R. J., and Nice, P. O. The experimental use of a programmed text in a medical school course. The Journal of Medical Education, 1962, 37 (8), 767-775.

Hebb, D. O. The Organization of Behavior. New York: Wiley, 1949.

Hebb, D. O. The semiautonomous process: its nature and nurture. American Psychologist, 1953, 18, 16-27.

Hively, W. Programming stimuli in matching to sample. Journal of Experimental Analysis of Behavior, 1962, 5 (3), 279-298.

Holland, J. G. Teaching machines: An application of principles from the laboratory. In W. I. Smith and J. W. Moore (eds.), Programmed Learning: Theory and Research. Princeton: Van Nostrand Co., Inc., 1962, 34-48.

Hughes, J. L. and McNamara, W. J. A comparative study of programmed and conventional instruction in industry. Journal of Applied Psychology, 1961, 45 (4), 225-231.

Iowa Tests of Basic Skills. Manual for Administrators, Supervisors and Counselors. Boston: Houghton-Mifflin, 1956.

Kilpatrick, F. P. Two processes in perceptual learning. Journal of Experimental Psychology, 1954, 47, 362-370.

Krumboltz, J. D., and Christal, R. E. Short-term practice effects in tests of spatial aptitude. Personality Guidance Journal, 1960, 38, 385-391.

Lambert, P., Miller, D., and Wiley, D. Experimental folklore and experimentation: The study of programmed learning in Wauwatosa Public Schools. Journal of Educational Research, 1962, 55, 485-494.

Lane, H. L. Experimentation in the language classroom: Guidlines and suggested procedures for the classroom teacher. Language Learning, 1962, 12 (2), 115-121.

Lorge, I., and Thorndike, R. L. The Lorge-Thorndike Intelligence Tests: General Manual. New York: Houghton Mifflin, 1955.

REFERENCES (Continued)

Lumsdaine, A. A., and Glaser, R. (eds.). *Teaching Machines and Programmed Learning*. Washington, D.C.: NEA, 1960.

Mager, R. F. On the sequencing of instructional content. *Psychological Reports*, 9, 1961, 405-413.

Margulies, S. The multiple-choice frame. *Programmed Instruction*, 1962, 1 (4), 4.

Markel, S. The lowest common denominator: A persistent problem in programming. *Programmed Instruction*, 1963, 2 (3), 4-5.

McNeil, J., and Keisler, E. Individual differences and effectiveness of auto-instruction at the primary grade level. *California Journal of Educational Research*, 1961, 12, 160-164.

McNemar, Q. *Psychological Statistics*. New York: Wiley, 1949.

Mendicino, Lorenzo. Mechanical reasoning and space perception: native capacity or experience. *Personality Guidance Journal*, 1958, 36, 335-338.

Miller, R. B. Analysis and specification of behavior for training. In. R. Glaser (ed.), *Training Research and Education*, Pittsburgh: University of Pittsburgh Press, 1962, 31-62.

Moore, W. J., and Smith, W. I. Knowledge of results in self-teaching spelling. *Psychological Reports*, 1961, 9, 717-726.

Myers, C. T. The effect of training or practice, or both, on scores on CEEB Spatial Relations Test, Form VACI. *ETS Research Bulletin*, March 16, 1951.

Myers, C. T. A note on a spatial relations pretest and posstest. *Educational and Psychological Measurement*, 1953, 13, 596-600.

Myers, C. T. The effects of training in mechanical drawing on spatial relations test scores as predictors of engineering drawing grades. *Research Bulletin RB-58-4*, ETS, March, 1958.

Naumann, T. A laboratory experience in programmed learning for students in educational psychology. *The Journal of Programmed Instruction*, 1962, 1 (1), 9-18.

Osgood, C. E. *Method and Theory in Experimental Psychology*. New York: Oxford University Press, 1953.

REFERENCES (Continued)

Pastore, N. Perceiving as innately determined. In L. D. Crow, Alice Crow (eds.), Readings in Human Learning. New York: David McKay Co., Inc., 229-235.

Ranucci, E. R. The Effect of the Study of Solid Geometry on Certain Aspects of Space Perception Abilities. Unpublished doctoral dissertation, Teachers College, Columbia University, 1952.

Reed, J., and Hayman, J. An experiment involving use of English 2600, an automated instruction text. Journal of Educational Research, 1962, 55, 476-484.

School Mathematics Study Group. Mathematics for Junior High School. Vol. I (Part 1) Ann Arbor: Photolithorpinted by Cushing-Malloy, Inc., 1960.

School Mathematics Study Group. Mathematics for Junior High School. Vol. I (Part 2) Ann Arbor: Photolithoprinted by Cushing-Malloy, Inc., 1960.

Shay, C. B. Relationship of intelligence to step size on a teaching machine program. Journal Educational Psychology, 1961, 52, 98-103.

Siegel, S. Nonparametric Statistics. New York: McGraw-Hill, 1956.

Silberman, H., Melarango, R., Coulson, J., and Estavan, D. Fixed sequence versus branching autoinstructional methods. Journal Educational Psychology, 1961, 52, 166-172.

Silvern, L. C., and Fuchs, F. H. Training Firemen Using an Aural-visual Teaching Machine. Los Angeles: Hughes Aircraft Company, 1962.

Skinner, B. F. Teaching machines. Science, 1958, 128, 969-977.

Skinner, B. F. Why we need teaching machines. Harvard Educational Review, 1961, 31, 377-398.

Stolurow, L. M. Teaching by Machine. Washington: U. S. Government Printing Office, 1961.

Thorndike, E. L.,and Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions, (I).(II) The estimation of magnitude; (III) Functions involving attention, observation, and discrimination. Psychological Review, 1901, 8, 247-261, 384-395, and 553-564.

Thurstone, L. L. Some primary abilities in visual thinking. The Psychometric Laboratory, University of Chicago, 1950, 59, 2-3.

REFERENCES (Concluded)

Thurstone, Thelma G., and Thurstone, L. L. *The Chicago Test of Primary Mental Abilities*. Chicago: SRA, 1958.

Thurstone, Thelma G. Primary mental abilities of children. In J. J. Jenkins and. D. G. Paterson (eds.), *Studies in Individual Differences*, New York: Appleton-Century-Crofts, 1961, 527-533.

Trow, W. C. *Educational Psychology* (2nd ed.). Boston: Houghton-Mifflin, 1950, 372-373.

Van Voorhis, W. R. *The Improvement of Space Perception Ability by Training*. Unpublished doctoral dissertation, Pennsylvania State College, 1941.

Walker, Helen and Lev, J. *Statistical Inference*. New York: Holt and Company, 1953.

Wendt, P. R. and Rust, G. Pictorial and performance frames in branching programmed instruction. *The Journal of Educational Research*, 1962, 55, 430-432.

Woolman, M. *Programming for Conceptual Understanding*. *Report to: The Communication Social Science Research Department of the Bell Telephone Laboratories*. Washington, D.C.: Institute of Educational Research, 1962.